FINAL REPORT

MASSACHUSETTS JUVENILE JUSTICE PAY FOR SUCCESS PROJECT

August 30, 2024[*]

EXECUTIVE SUMMARY

BRIEF OVERVIEW OF PROJECT: Among the first of its kind, the Massachusetts Juvenile Justice Pay for Success ("PFS") Project was an experiment in a new way to finance and procure vital social services. The Commonwealth of Massachusetts, Roca Inc., and Third Sector Capital Partners, Inc., in cooperation with funders Goldman Sachs, The Kresge Foundation, Living Cities, Laura and John Arnold Foundation (now Arnold Ventures), New Profit Inc., and The Boston Foundation, launched the initiative to test this innovative model while working to reduce recidivism and improve employment outcomes for young men at high risk of re-offending in the Boston, Chelsea and Springfield, Massachusetts areas.

After nearly two years of planning and negotiation, the project launched in early 2014. According to the agreements:

- **The Commonwealth of Massachusetts** would pay up to $28 million to the project based on evidence that a provider reduced incarceration, increased employment, and prepared young people at high risk of incarceration to be economically independent.
- **Roca,** a nationally known nonprofit with over 20 years of experience in the field, would serve up to 1,320 high-risk young men in the Boston, Chelsea and Springfield, Massachusetts regions by providing its evidence-based intervention.
- **An independent evaluator** (whose identity changed several times over the life of the project but was ultimately Akiva Liberman, Ph.D.) would administer a randomized control trial (RCT) to assign 17- to 24-year-old young men who were at high risk of recidivating and were on probation, in the custody of the Department of Youth Services (DYS), or leaving an adult custodial institution[1] to either (i) a treatment group of young people whom Roca would try to serve or (ii) to a control group, who would receive no special services. The Evaluator would then determine whether the treatment group spent fewer days in prison and more quarters employed than the control group. Additionally, under certain circumstances, the Evaluator would combine the RCT with a Difference in Differences (DID) study to make the final outcomes determination.
- **A group of funders -- Goldman Sachs, The Kresge Foundation, Living Cities, Laura and John Arnold Foundation (now Arnold Ventures), New Profit Inc., and The Boston Foundation** – provided $16 million in loans and grants to pay for Roca's services and the evaluation and other project costs. Depending on whether and the extent that the project decreased days of incarceration and increased employment, the lenders would receive interest, repayment of

[*] This report has been prepared by Roca, Inc., the Commonwealth of Massachusetts, and Third Sector Capital Partners, Inc.

[1] This was the final study population. The parties agreed to expand beyond the original definition, which included only people on probation or released from DYS, to partially compensate for low and slow referrals.

principal, and a small return, and grantors would be able to redeploy their grants together with a small return. Roca also provided financial support to the project by deferring $3.5 million in fees with payment dependent on outcomes. The Commonwealth ultimately provided nearly $12 million to the project: $6.1 million in "job readiness payments'' in exchange for Roca preparing young people to enter the workforce, $1.3 million to compensate for the low number of referrals and $4.5 million related to the project extension.[2]

- **Third Sector Capital Partners, Inc**., a nonprofit technical assistance provider, would serve as the project manager and its supporting organization, **Youth Services Inc. ("YSI"),** would be the financial intermediary, serving as the borrower and grantee and directing payments to and from the parties.

PROJECT EXECUTION:

The project encountered challenges from the beginning. There were not enough high-risk young people in the system to randomly assign to Roca or to the control group. And those referred to Roca often had bad addresses, making it impossible for Roca to find and engage them in services. As a result of these challenges, the parties repeatedly renegotiated their agreements to execute on the envisioned project and evaluation as well as possible. Major contract amendments that included changes to the funding structure were executed in 2016 and 2020, and the project was extended by three years and three months, to March 31, 2024, to allow more time to observe results.

Additionally, in 2016 Roca restructured its Intervention Model to integrate a version of cognitive behavioral theory (CBT) it developed with Massachusetts General Hospital – Rewire CBT -- helping young people develop the tools necessary to manage the chronic and ongoing trauma they have experienced and learn the skills they need to achieve sustained behavior change. Unfortunately, however, because of the timing of this project, 83% of participants were not fully trained and coached in Rewire CBT, and as such, they did not receive the current version of Roca's Intervention Model.

There were also at least two major external challenges to executing the project as originally envisioned: (1) a significant change in sentencing law and practices in the Commonwealth and (2) the Covid epidemic of 2020.

Nevertheless, the project proceeded, and Roca served 1082 young people over ten years, and the Evaluator issued a final report in June 2024

2 The job readiness payments were part of the original agreement between the parties, and the $1.3 million was negotiated as part of the 2016 contract amendment when the low number of referrals and resulting reduction in the statistical power of the evaluation became clear. The $4.5 million related to the project extension was added in 2020 when the parties agreed to continue the project, primarily to allow the evaluation to collect additional data. All these payments were subject to the $28 million cap on payments from the Commonwealth.

OUTCOMES:

The project had many positive results:

- A complex experiment in social financing and government procurement was executed over a ten-year period and the parties worked together collaboratively to overcome numerous challenges.
- A total of 1082 young people benefitted from an opportunity to learn life skills that would help them address their trauma, make choices that would keep them alive and out of prison and get jobs.
- Roca received over $23.8 million in fees over 10 years – —funding that has contributed meaningfully to the organization's growth and development. Roca's enhancements to its Intervention Model have benefitted thousands of young people at the center of urban violence across the Commonwealth who achieved life-changing outcomes outside of the PFS project.
- The evaluation, including the RCT and the DID study, which was used, was executed as described in the Evaluation Plan, which is included an appendix the Final Evaluation Report (Attachment A).
- Finally, the parties, and interested colleagues around the country, learned valuable lessons about working with Massachusetts criminal justice data and how to construct PFS projects.

The evaluation for this project produced statistically inconclusive results and will not cause the Commonwealth to make any success payments. The final report of the Evaluator (Exhibit A) determined that the RCT showed that young people in the treatment group on average were incarcerated for 43 days more than those in the control group (with the 95% confidence interval ranging from 21 fewer days to 108 more) and were employed for 0.12 fewer quarters (with the 95% confidence interval ranging from 0.43 more quarters to 0.66 fewer). The DID estimate, which looked at the same group of people receiving the Roca intervention but compared them to different control groups, found that Roca decreased incarceration by 17 days and increased employment by 0.7 quarters per person.[3] This study, however, was also statistically insignificant, and we recognize that RCT's are generally considered the "gold standard" of evaluation. At the same, the different directions of the RCT and the DID show how sensitive the estimates of impact are to the method used. [4]

The Commonwealth did make $7.4 million in success payments for the increased job readiness of young people served – all of which went to Roca -- and to address the lack of precision caused by the shortfall in referrals. The Commonwealth made an additional $4.5 million in Retention and Completion payments and Project Extension Expense payments that allowed Roca services and the evaluation to continue from January 2020 through March 2024.

---

[3] This discussion uses the intent to treat (ITT) estimates of both the RCT and the DID studies. The Evaluator also performed instrumental variable estimates (IVE) to account both for the fact that only a portion of the treatment group received Roca's services and the possibility that Roca would unknowingly enroll people in the control group in the ordinary course of its operations unrelated to this project. The final project evaluation combined the two IVE estimates to create "the final backstop evaluation."

[4] When the Evaluator combined the two studies for the final evaluation, it showed that the project increased incarceration and decreased employment.

When considering the inconclusive results of the evaluation, it is important to consider the challenges faced by the project as well as the fact that the evaluation was significantly less predictive than originally planned. **In the words of Professor Jeffrey Liebman of the Harvard Kennedy School (HKS) and the Commonwealth's advisor: this evaluation "ended up extremely underpowered and only marginally informative."**

The parties understood from the start that there might be challenges to the evaluation and that the results might not be definitive. As a result, the contract among and between the Commonwealth, Roca and YSI provides: "[The evaluation is] intended to provide a contractual mechanism for payment. [It] shall [not] be used by the Commonwealth, nor [is it] intended for use by any other person or entity, to characterize the impact of the Roca Model in any other context."

In addition to the Evaluator's DID study, an additional analyses sheds light on the extent to which the estimated impacts of this program are sensitive to the approach used to make the estimate. Roca's independent evaluator, Abt Global, conducted a corresponding study looking at all the young people served by Roca (both as part of this project and not). According to that report, "Although Roca serves only the most high-risk individuals, its participants have lower rates of reincarceration than are seen for Massachusetts in the same cohort . . .. Roca participants consistently have lower rates of reincarceration than the state sample at one, two, and three years. Among the Roca participants at highest risk for reincarceration — those who have previously been incarcerated for a violent offense — rates of reincarceration for a violent offense are also lower than the Massachusetts all-counties total reincarceration rates."

The final report of the Evaluator, Akiva Liberman, Ph.D. is Attachment A to this memo and Abt's report is included as Attachment B.

FINAL REPORT
MASSACHUSETTS JUVENILE JUSTICE PAY FOR SUCCESS PROJECT
August 26, 2024

TABLE OF CONTENTS

MASSACHUSETTS JUVENILE JUSTICE PAY FOR SUCCESS PROJECT -- FINAL REPORT*

INTRODUCTION

Among the first of its kind, the Massachusetts Juvenile Justice Pay for Success ("PFS") Project was an experiment in a new way to finance and procure vital social services. The Commonwealth of Massachusetts, Roca, Inc., and Third Sector Capital Partners, Inc., in cooperation with funders Goldman Sachs, The Kresge Foundation, Living Cities, Laura and John Arnold Foundation (now Arnold Ventures), New Profit Inc., and The Boston Foundation, launched the initiative to test this innovative model while working to reduce recidivism and improve employment outcomes for young men at high risk of re-offending in the Boston, Chelsea and Springfield, Massachusetts areas.

The government often focuses on reacting to a problem rather than preventing it. This is frequently more expensive, both in terms of tax dollars deployed and the costs borne by our residents, families, and communities. And on those occasions when the government does fund preventative services, much of the time, it pays based on the number of people served and often is not able to measure whether services achieve desired results. What is more, not only does the government far too often spend taxpayer money without full understanding of the results, the provider of services does not have access to the government (or administrative) data that would allow them to understand how they are doing and to course correct. Finally, the government procures services annually, destabilizing providers and preventing them from engaging in long-term planning.

The PFS model was conceived to address these problems. It often has 5 components:

1. **Government procures a social service and articulates the outcome it is seeking, for instance, reduced recidivism or increased employment, within a described population, and offers to pay for the services if the outcome is achieved instead of based on number of people served.** Often, if the services are successful in preventing the underlying problem, they will be less costly than trying to fix it after the fact. For example, a program that was highly successful at keeping people out of prison could be cheaper than incarcerating them.
2. **A service provider contracts with the government to provide the services and be paid based on outcomes (and often other proxy indicators of success).**
3. **An independent evaluator uses government data to determine whether the provider has achieved the outcomes, and the government makes the required payments.** The evaluator shares this data with the service provider throughout the project.
4. Often, the provider does not have the money to pay for its services while the evaluator determines whether outcomes were achieved or the capacity to take on the risk that government will not pay. **As a result, socially minded funders will put up the money to pay for services with the hope of getting their money back and some level of return and take on the financial risk of this model.**

---

* This report has been prepared by Roca, Inc., the Commonwealth of Massachusetts, and Third Sector Capital Partners, Inc.

5. **Finally, because PFS projects can be complex, the parties often engage a project manager.**

Recognizing the potential of the PFS model, Massachusetts was the first state in the nation to issue a competitive procurement for services using this structure on January 18, 2012. After months of planning and negotiation this project launched in early 2014. According to the agreements:[5]

- **The Commonwealth of Massachusetts** would pay up to $28 million to the project based on evidence that a provider reduced incarceration, increased employment, and prepared young people at high risk of incarceration to be economically independent. The Commonwealth ultimately provided nearly $12 million to the project: $6.1 million in "job readiness payments" in exchange for Roca preparing young people to enter the work force, $1.3 million to compensate for the low number of referrals, and $4.5 million related to the project extension.[6]
- **Roca** would be the service provider serving up to 1,320 high-risk young men over 9 years in the Boston, Chelsea and Springfield, Massachusetts regions by providing its evidence-based intervention. Roca is a nonprofit that has delivered its high impact intervention to young people in Massachusetts areas for 36 years. The program had a proven track record of reducing incarceration rates among the highest risk individuals.
- **An independent evaluator** (the Evaluator), which began at start-up with an evaluation firm called **Sibalytics** and moved over the course of the project to **The Urban Institute**, then **Child Trends,** and finally **Akiva Liberman, Ph.D.** (who was the lead evaluator at The Urban Institute and Child Trends), would administer a randomized control trial (RCT). According to the evaluation plan, the Evaluator would randomly assign 17- to 24-year-old young men who were at high risk of recidivating and were on probation, in the custody of DYS, or leaving an adult custodial institution[7] to either a "treatment group" of young people, whom Roca would try to serve, or to a control group, who would receive no special services. Additionally, under certain circumstances, the Evaluator would combine the RCT with a Difference in Differences (DID) study to make the final outcomes determination. The Evaluator would then determine whether the treatment group spent fewer days in prison and more quarters employed than the control group.
- **The Funders:**
  - **Goldman Sachs** would provide $8 million in senior loan financing.
  - **The Kresge Foundation and Living Cities** together would provide $2.66 million ($1.33 million each) in junior loan financing.
  - **Laura and John Arnold Foundation (now Arnold Ventures), New Profit, and The Boston Foundation** together would provide $5.45 million in grants ($3.34 million from Laura

---

5 For simplicity's sake, the figures in following section are those arrived at after the many minor and two significant amendments to the project agreements. Details of the changes are set out in Attachment C.

6 The job readiness payments were part of the original agreement between the parties, and the $1.3 million was negotiated as part of the 2016 contract amendment when the low number of referrals and resulting reduction in the statistical power of the evaluation became clear. The $4.5 million related to the project extension was added in 2020 when the parties agreed to continue the project, primarily to allow the evaluation to collect additional data. All these payments were subject to the $28 million cap on payments from the Commonwealth.

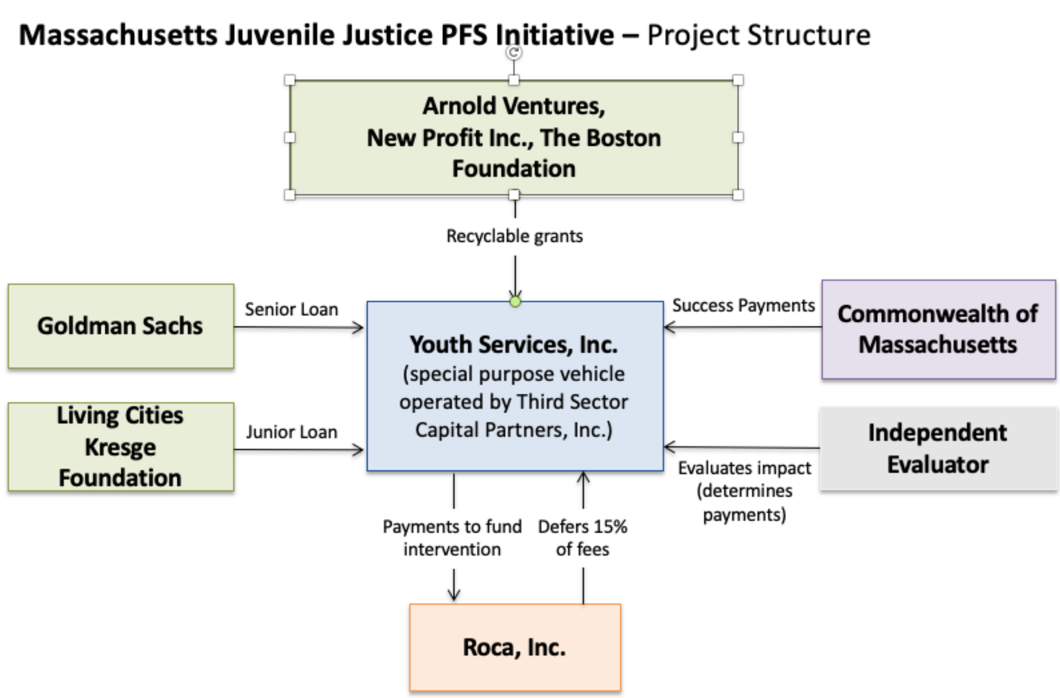
7 This was the final study population. The parties agreed to expand beyond the original definition to partially compensate for the "low and slow" referrals.

and John Arnold Foundation, $1.81 million from New Profit, and $300,000 from The Boston Foundation).

    - **Laura and John Arnold Foundation** (now Arnold Ventures) would provide an additional $3.6 million to support the costs associated with the final project evaluation, allowing outcomes for all young people in the project to be observed.[8]

- Depending on whether and to what extent recidivism and employment outcomes were achieved, the lenders would receive interest, repayment of principal, and a small return, and grantors would be able to redeploy their grants and a small return. Roca also provided financial support to the project by deferring $3.5 million in fees with payment dependent on outcomes.
- **Third Sector Capital Partners, Inc**., a nonprofit technical assistance provider, would serve as the project manager and its supporting organization, **Youth Services Inc. (YSI)** would be the financial intermediary, serving as the borrower and grantee and directing payments to and from the parties.
- **The Harvard Kennedy School (HKS) Government Performance Lab (originally, the Social Impact Bond Technical Assistance Lab)** assisted Massachusetts in developing the procurement, designing the data analysis strategy for this project, and provided ongoing advice and assistance.

Figure 1: The MAJJ PFS Project



8 Laura and John Arnold Foundation, now Arnold Ventures, funded this work through a direct grant to Roca in October of 2016.

Included as Attachment A to this report is the final report of the independent project evaluator and as Attachment B: additional analysis from Roca's independent evaluator, Abt Global. This report summarizes and contextualizes the conclusions of those two reports.

PROJECT SERVICES

Roca helps young men who are or have been involved with the Commonwealth's justice system break the cycle of reoffending by increasing workforce participation and job readiness. Roca does this through its proven Intervention Model, which connects very high-risk youth to each other and adults through intensive relationships and uses targeted life skills, education, and employment programming to support young people in developing the skills necessary to reduce violence and create positive behavioral changes. Roca's four-year model—which consists of two years of intensive engagement and two years of follow-up—includes four basic elements: relentless outreach to young men by Roca staff; intensive case management; life skills, educational, prevocational and employment programming; and work opportunities with community partners.

In 2016, after a preliminary pilot, Roca restructured its Intervention Model and fully grounded it in the delivery of Rewire CBT, a tool the organization believes is critical to the current success of its young people, helping them develop the tools necessary to manage the chronic and ongoing trauma they have experienced. Unfortunately, because of timing of this project, 83% of participants were not fully trained and coached in Rewire CBT; they did not receive Roca's Intervention Model as it is currently delivered.[9]

THE PROJECT EVALUATION

The parties agreed that the core project evaluation would be a randomized control trial. Under the original evaluation plan, the Commonwealth would refer 1,821 young men aged 17-24 who were in the Boston, Chelsea, or Springfield areas and (i) leaving the juvenile justice system or (ii) were involved in the probation system to the Evaluator. The Evaluator would then screen those referred based on several factors, primary among them was being at high-risk for re-incarceration. Based on a historical analysis by the HKS Government Performance Lab, all the individuals leaving the juvenile justice system were deemed to be high-risk. The Evaluator determined who, among those leaving state or county incarceration or otherwise on probation, were high-risk by relying on the risk score used by the Department of Probation. The Evaluator then randomly assigned eligible young men to one of two

[9] Roca's CBT-based intervention and approach to youth work, ***Rewire CBT – Skills for Living (Rewire)***, developed with Massachusetts General Hospital, can be taught, practiced, and mastered in the street or in community settings. It was designed to overcome traditional barriers to access, helping front-line staff reach high-risk young people who tend to refuse interventions in clinical settings. **Rewire** is simple, mobile, and relatable to the lives of young people involved in violence. Roca's approach is grounded in the teaching and modeling of core CBT skills in real time to build young people's: **emotional regulation** (their ability to manage their emotions and slow down the fight, flight or freeze response), **behavior activation** (ability to do something different and move out of bottom-brain living), and **cognitive flexibility** (ability to think differently and make intentional choices.) Rewire gives people a profoundly empowering tool to access their pre-frontal cortex and reclaim the power of choice.

groups: "the treatment group," who were referred to Roca for services, and "the control group". The plan assumed that Roca would enroll 1275 (70%) of those the Evaluator referred to them.

Roca was allowed to enroll young men whom the Evaluator did not refer to them but found their way to Roca in their own – Roca did not have to deny services to anyone and served many young men outside of the project over the past ten years. And the evaluation planned for this cross-over.

It is important to note that young people who were referred to Roca were not mandated to participate in Roca programming. Rather, it was Roca's job to relentlessly connect with these young people and provide them services regardless of each young person's inherent willingness and motivation to participate. Roca was chosen for this project because of its historical success with populations that others find difficult to engage.

The Commonwealth's advisor, Harvard Kennedy School Professor Jeffrey Liebman explains:

> The statistical power calculations that were undertaken prior to the launch of the program anticipated a minimum detectable effect (MDE) on incarceration of 30 percent for the instrumental variable results. These calculations implied that if the true effect of Roca's intervention was to reduce incarceration by 30 percent, there would be at least an 80 percent chance that the evaluation would find that the intervention reduced incarceration by a statistically significant amount (at a 95 percent confidence level). These calculations also implied that the 95 percent confidence interval on the experimental results was expected to have a width of approximately 40 percentage points; if the evaluation were to find a point estimate of 30 percent, the 95-percent confidence interval would have extended from 10 percent to 50 percent. These calculations also implied that if the true effect of the Roca model was that it reduces incarceration by the 40 percent that was anticipated in the financial model – there would be only a negligible (2.5 percentage point) chance of the evaluation finding an estimated effect below 20 percent. And it meant that if the true effect of Roca was that it reduced recidivism by 20 percent, there would be an 84 percent chance that the evaluation would estimate an impact of at least 10 percent.

The plan also anticipated that the statistical power would be reduced if too many young people randomized to the control group found their way to Roca on their own. As such, the original plan included a "Backstop Methodology" where, if a certain percentage of the control group enrolled in Roca, a quasi-experimental DID evaluation would be added to the RCT before the final evaluation.

The original evaluation plan anticipated that the final randomization would be in September 2016, and the final report would be issued at the end of 2019. The final payments from the Commonwealth would follow thereafter.

Knowing that there might be challenges to this evaluation, some of which are discussed below, the parties acknowledged from the outset that the evaluation was designed for the specific purpose of this PFS project and not as a full evaluation of Roca's intervention model. The Pay for Success Contract that the parties approved provided:

> For avoidance of doubt, the RCT and Backstop Methodology are intended to provide a contractual mechanism for payment. Neither shall be used by the Commonwealth, nor

are they intended for use by any other person or entity, to characterize the impact of the Roca Model in any other context.

POTENTIAL PAYMENTS

The parties originally contracted for three types of payments from the Commonwealth:

1. **Payments for decreases in incarceration** represented the majority of the potential success payments and were based on a graduated payment schedule where the Commonwealth would pay increasing amounts for each day that participants avoid incarceration as compared to similar young men who are not in the program.

2. **Payments for increases in employment** were set at $750 for each participant in each quarter that a Roca participant is employed as compared to similar young men who are not in the program.

3. **Payments for increases in job readiness** were $1000 for each participant in each calendar quarter that a Roca participant engaged with a Roca youth worker nine or more times, with each engagement helping young men address barriers to employment and move toward economic independence.

The first two types of payments would be made based on the RCT (and if needed, the Backstop Methodology). The third was based on reports from Roca that would be validated by an independent third party.

REFERRALS AND CONTRACT AMENDMENTS

Throughout the project, a range of issues gradually emerged related to the number of young men referred and their suitability to the project. For example, the project had difficulty in its first year finalizing the referral process, and the project consistently struggled to identify an adequate number of young people to randomize and refer to the project.

As a result, Roca enrolled many fewer young people and did so at a later point in the project than expected, and the control group was smaller than expected. Further, Roca was able to enroll and treat a much lower percentage of those referred to it than expected. (An estimated 1/3 of referrals were sent with bad contact information and 1/3 were too low risk by Roca standards to be served.)

As a result of these challenges, the statistical power of the evaluation was much lower than expected. Further, as Roca was paid based on the number of young people enrolled, the low number of referrals created a financial risk for Roca. The parties could have mutually agreed to end the project early or possibly, could have put the Commonwealth in default, requiring the Commonwealth to repay all costs to date. Instead, on multiple occasions, all parties agreed to amend the agreements while maintaining the structure of the evaluation and the project. While each party no doubt had its own changing motivations for doing so, they were all united in a commitment to testing the PFS model and in a belief that the Roca model would help young people and that continuing the project would benefit them and the Commonwealth as a whole.

Across 5 side letters and two major contract revisions and amendments, there were many changes (see Attachment C for details) to the agreements among the parties. The most important were:

- The project would draw referrals not only from Probation and DYS but also from young people on parole or leaving state or county correctional facilities. These new referral sources did not use the same risk score as Probation. As a result, the HKS Government Performance Lab developed a risk score for referrals from these sources that was used only for this project.
- The initial plan was for Roca to receive as many as 1,821 referrals and enroll 70%, or 1275 young people. In practice, the evaluation referred 1186 people to Roca, and Roca enrolled 37.6%, or 446 of them, reducing statistical power of the evaluation and the reliability of its results. That is, lower statistical power increases the chanced of underestimating, overestimating, or even flipping the direction of the true effect of Roca's intervention. Please see Attachment C for more information regarding changes in referral expectations over time.
- Roca was allowed to "self-recruit" program participants who would be paid for by the project but not included in the evaluation.
- The Commonwealth would make additional payments to the project (all of which went to Roca) to compensate for the reduction in referrals and resulting decrease in the statistical power of the evaluation.
- The Backstop Methodology would be used not only if the control group was too "contaminated" by Roca "self-recruits", as originally planned for, but also if the number of referrals who qualified for participation in Roca did not reach a certain threshold.
- Referrals would extend from quarter 12 of the project to quarter 21, and the final evaluation would be due no later than March 31, 2024, instead of December 2019. The Commonwealth agreed to pay for Roca's services and project costs during this extension.

PROJECT OUTCOMES

The final report of the project evaluation determined that the Commonwealth would not be required to make any payments tied to the incarceration or employment outcomes of the young people assigned to Roca. As a result, while the lenders were paid over $1.5 million in interest, the project did not repay any of the $10.7 million in principal, and the loans were cancelled; the grantors were not able to redeploy any of their $5.45 million in "recyclable grants"; and Roca did not recoup any of its $3.5 million in deferred fees.

The table below depicts the results of the various evaluations the Evaluator performed. To facilitate understanding and highlight the broad range of outcomes that this evaluation could support, the estimates that are "beneficial" to Roca – that is, a decrease in incarceration and an increase in employment -- are highlighted in green. The final estimate is a combination of the IVE of the RCT's (80%) and the DID (20%). [10]

[10] For more details around the evaluation, see Attachment A, the Final Evaluation Report.

Summary of Evaluation Results		
	Average Difference per Person in Days of Incarceration Between Treatment and Control	Average Difference Per Person in Quarters of Employment Between Treatment and Control
RCT -- Intent to Treat		
Low End of 95% Confidence Interval	-21	-0.7
Estimate	43	-0.1
High End of 95% Confidence Interval	108	0.4
RCT --Instrumental Variable		
Low End of 95% Confidence Interval	-70	-2.2
Estimate	151	-0.4
High End of 95% Confidence Interval	372	1.4
DID -- Intent to Treat		
Low End of 95% Confidence Interval	-145	-0.8
Estimate	-17	0.3
High End of 95% Confidence Interval	111	1.25
DID - Instrumental Variable		
Low End of 95% Confidence Interval	-379	-0.7
Estimate	-44	0.7
High End of 95% Confidence Interval	291	3.3
Final Backstop Estimate		
Estimate	112	-0.2

Here again, four things must be emphasized about the evaluation:

1. It is a contractual mechanism for payment – not an evaluation of the Roca Model as a whole. The Pay for Success Contract that all parties signed off on provided: "Neither [the RCT [nor] Backstop Methodology] shall be used by the Commonwealth, nor are they intended for use by any other person or entity, to characterize the impact of the Roca Model in any other context."
2. The project did not assess the current Roca Model, a model that has added a critical component designed to address trauma and the behavior and thinking that are products of trauma.
3. Two other analyses using different methodologies, as well as the DID, all of which included some or all of the same young people as the RCT, illustrate the extent to which conclusions about the impact of this project are sensitive to alternative methodological assumptions and found that Roca reduced incarceration. None of these evaluations reached a conclusion with statistical certainty.
4. The project encountered a number of complications that reduced the statistical power of the evaluation. As described by Professor Liebman: *"The actual experiment ended up extremely underpowered and only marginally informative."*

SUPPLEMENTARY EVALUATION FINDINGS

Abt Global assessed outcomes for all of the young people who enrolled in Roca in 2017, both through the project and outside of the project, and compared their incarceration rates to the state's average. While many of these young people received Roca's Rewire CBT, that was just being rolled out for this cohort so not all young people received the full Rewire CBT training and coaching.[11]

According to Abt Global:

> Although Roca serves only the most high-risk individuals, its participants have lower rates of reincarceration than are seen for Massachusetts in the same cohort (2017) . . .. [W]e see that Roca participants consistently have lower rates of reincarceration than the state sample at one, two, and three years. Among the Roca participants at highest risk for reincarceration —those who have previously been incarcerated for a violent offense — rates of reincarceration for a violent offense are also lower than the Massachusetts all-counties total reincarceration rates. **Roca's incarceration rate for these cohorts was 23.8% at one year (compared to 27.5% statewide); 34.4% at two years (compared to 42.9% statewide); and 37.7% at three years, (compared to 50.1% statewide.)**

Complete details on this analysis can be found in Attachment B.

QUESTIONS ABOUT THE PROJECT

HOW DID THE FINAL OUTCOMES DIFFER FROM THE EXPECTED OUTCOME OF THE PROJECT?

Based on HKS Government Performance Lab's historical analysis of high-risk young people in the Commonwealth, the project was built on a "base case" that assumed that approximately 64% of people in the control group would be incarcerated, each for an average of 840 days, and that Roca would reduce incarceration by 40% --to an average of 504 days.

Instead, 38% of people in the control group were incarcerated for an average of 329 days per person The young people in Roca were incarcerated for 367 days on average. This is an increase of 12% as compared to the control group but a 56% reduction compared to the historical baseline.[12]

---

11 Of those served in 2017 by Roca, 93% received training on at least one CBT skill (210 out of 222 young people.) Of those who received Rewire CBT, the average participant received 52 Rewire CBT contacts between 2017 and 2020.

12 These numbers are descriptive – they do not account for any control variables. The independent evaluator explains how that may skew these figures: "Especially in early estimates, the uncontrolled comparisons tend to be systematically biased toward larger outcomes for the treatment group. This is the result of more individuals having been intentionally randomized to treatment in early quarters in the project (see Appendix A), and therefore having longer observed periods of follow-up, which then gave them greater opportunity to accrue outcomes, whether detrimental (i.e., being sentenced to incarceration) or beneficial (i.e., employment)."

WHAT HAPPENED?

While it is difficult to determine why the results of ROCA's intervention did not match expectations, a few key factors are pertinent to the analysis:

1. The RCT was executed – despite real challenges – as contracted for.

2. Incarceration across the Commonwealth is much lower than it was when the project began in 2014.

For example:

i) The prison population has consistently fallen since its peak in 2012.
ii) It dropped by 48% from 2012 to 2023.
iii) The three-year recidivism rate of the overall prison population released in in 2018, regardless of gender, age or risk level, was 29%, the lowest on record since the state began tracking that statistic in 1995.[13]

This drop in incarceration is likely in some part the result of changes in practice and legislation[14] and is plainly a good thing for those young people who are no longer being needlessly incarcerated and for the Commonwealth as a whole. It did, however, change the probability of the project being financially successful.

We cannot specifically tie this project to that drop. At the same time, we cannot rule out the very real possibility that the project and Roca's work in general contributed to the statewide reforms and efforts that led to this drop.

3. There is a discrepancy between the statewide reincarceration rate and the reincarceration rate observed in the project.

As described above, before adjustment for covariates, 38% of the control group -- a group assessed to be "high risk" --were incarcerated over five years, including people released in 2017, 2018 and 2019. But according to the Massachusetts Executive Office of Public Safety and Security (EOPSS), the *three-year reincarceration rate* for all 18- to 24-year-old young men released from county correctional facilities, regardless of risk level, was 50.1% for 2017 releases; 42.7% for 2018 releases and 40.6% for 2019 releases.[15] Thus the five-year reincarceration rate for the high risk people in the control group was lower than the three-year rate for all people released, low and high risk. This is incongruous; one would

---

13 https://www.mass.gov/lists/prison-population-trends; https://www.mass.gov/doc/frequently-asked-questions-january-2024/download

14 In 2018, the governor signed a sweeping criminal justice reform bill into law. The landmark bill aimed to develop a more equitable system that supports the state's youngest and most vulnerable residents, reduces recidivism, increases judicial discretion, and enhances public safety. Highlights of the bill that likely impacted this project include: (1) the elimination of mandatory and statutory minimum sentences for many low-level, non-violent drug offenses, and (2) the requirement of district attorneys to create pre-arraignment diversion programs.

15 Massachusetts Executive Office of Public Safety and Security, Cross Tracking System – Recidivism – Query Model, found at https://www.mass.gov/info-details/cross-tracking-system-recidivism-query-model#downloads-

expect high risk probationers observed over five years to be reincarcerated more frequently than a mixed group of people observed over three years.

It is not clear what caused this discrepancy. One possible explanation is that the project missed some reincarcerations because of a delay in randomization: as described in the Final Evaluation Report, the median time between when someone was released from incarceration until they were randomized was 196 days.[16] This likely happened because of a gap between a person starting probation and Probation assigning a risk score to that young person. Neither the parties nor the Evaluation Plan anticipated this gap. Without a risk score, however, a young person could not be randomized and assigned to the treatment group or the control group. We cannot know whether -- and if so, how -- this unexpected complication might have impacted the project. It almost certainly changed the population of people in the study and their recidivism rates and how Roca interacted with its young people.

4. The statistical power of the final estimate was far less than what the parties originally expected, making the evaluation of extremely limited utility other than as a mechanism for payment.

Professor Jeffrey Liebman, who has advised the Commonwealth since the project's earliest days writes:

> The statistical power calculations that were undertaken prior to the launch of the program anticipated a minimum detectable effect (MDE) on incarceration of 30 percent for the instrumental variable results. These calculations implied that if the true effect of Roca's intervention was to reduce incarceration by 30 percent, there would be an 80 percent chance that the evaluation would find that the intervention reduced incarceration by a statistically significant amount (at a 95 percent confidence level). These calculations also implied that the 95 percent confidence interval on the experimental results was expected to have a width of approximately 40 percentage points; if the evaluation were to find a point estimate of 30 percent, the 95-percent confidence interval would have extended from 10 percent to 50 percent. These calculations also implied that if the true effect of the Roca model was that it reduces incarceration by the 40 percent that was anticipated in the financial model – there would be only a negligible (2.5 percentage point) chance of the evaluation finding an estimated effect below 20 percent. And it meant that if the true effect of Roca was that it reduced recidivism by 20 percent, there would be an 84 percent chance that the evaluation would estimate an impact of at least 10 percent.
>
> The original power calculations assumed that 1821 individuals would be referred to Roca and that Roca would be able to enroll 70 percent of them (1275) for services. The original calculations also assumed that 781 individuals would be assigned to the control group and that 20 percent of the control group would end up receiving services from Roca – so there would be an experimental contrast in terms of exposure to Roca of 50 percentage points. Finally, the original calculations assumed a coefficient of variation for incarceration days of approximately 1.2. In practice, only 1186 people were referred to Roca, and Roca enrolled 37.6 percent of them (446 individuals). The actual control group was 626, 8.6 percent of whom received services from Roca. So, the number of people served by Roca was much lower than expected, and the experimental contrast between the treatment group and control was only 29

[16] The Evaluator was able to obtain complete data on time to randomization for 773 of the 1818 cases in the RCT. The mean time to randomization was 317 days.

percentage points. The coefficient of variation was also much larger than predicted – approximately 1.9.

> **The actual experiment ended up extremely underpowered and only marginally informative**. [emphasis added] Plugging in the actual program parameters into the original statistical model, the minimum detectable effect is a 91 percent reduction in incarceration. The 95 percent confidence intervals are plus or minus 65 percent – implying that if a point estimate of zero were estimated, we would not be able to rule out a 65 percent decrease in incarceration or a 65 percent increase in incarceration. These theoretical calculations match quite closely the actual estimates in this report. The 95 percent confidence interval on the incarceration IV estimate is plus or minus 67 percent of the control group mean. And the actual 95 percent confidence interval for the incarceration IV estimate extends from a 21 percent reduction in incarceration days to a 113 percent increase in incarceration days.

5. The COVID Pandemic likely had some unknowable impact.

We cannot know the full implications of the COVID pandemic on the project. We do know, however, that COVID closed courts for month and subsequently resulted in long delays in trials and sentencing. Additionally, crime and reincarceration rates fell.

The Independent Evaluator writes:

> COVID-19 affected criminal justice processes in complex ways, especially in ways designed to reduce transmission of the virus, such as the use of community-based rather than incarcerative sentences, less use of pretrial detention, and/or changes in community supervision and probation requirements. Employment and employment opportunities were also dramatically affected. Simultaneously, criminal behavior changed in complex ways in many U.S. cities, with some reporting declines in property crime, increases in some violent crimes, and surges in drug overdoses; some of these changes persisted even as the most acute phase of COVID-19 passed. Finally, Roca has reported that program services were delivered during COVID but that they were adjusted from the traditional in-person services delivered in the community to virtual services, in order to allow for social distancing. In short, COVID-19 has changed the program and likely affected the outcomes being monitored for the evaluation. Thus, we may expect COVID-19 to have affected outcomes, although we have no *a priori* expectations for how COVID-19 would *differentially* affect outcomes of people in the treatment versus control groups.

6. The Roca Intervention was NOT Consistent Throughout the Project.

Unlike traditional evaluations in which a program model is locked in from the start and great effort is made to maintain program fidelity so that it is clear exactly what is being evaluated, the PFS model envisioned that the provider would receive high frequency feedback on performance and improve its program model during the course of the project. During this project, Roca made one comprehensive, critical change that is important to note.

In 2016, after a preliminary pilot, Roca's Intervention Model was restructured and fully grounded in the delivery of Rewire CBT, a tool the organization believes is critical to the current success of their young people, helping them develop the tools necessary to manage the chronic and ongoing trauma they have experienced. Unfortunately, however, because of timing of this project, 83% of program participants in the MA Juvenile Justice Pay for Success Project were not fully trained and coached in Rewire CBT nor did they receive Roca's Intervention Model as it is currently delivered.

WHERE DID THE FUNDERS' DOLLARS GO?

Providing Roca services to the 1082 young people, running the evaluation, validating the results of the evaluation and project management cost $29.9 million. Additionally, the project paid funders $1.7 million in interest.

Project costs were covered by $16.1 million in loans and grants from the funders, $3.5 million in expenses advanced by Roca, and $12 million in payments from the Commonwealth. These included payments based on increase in job readiness among the young men assigned to Roca, to compensate for the lack of referrals, to add young people to the evaluation, and to cover the costs of extending the project.

WHAT ARE SOME OF THE LESSONS PROJECT PARTIES HAVE LEARNED?

1) **Always run a pilot.** The first referrals to the project happened in the first quarter of 2014 and were part of the project's official evaluation. While it immediately became apparent that there were not as many eligible young men as planned, any modifications required changes to the project contracts and evaluation plan. It was a mistake not to have "practice runs" where changes would have been less consequential.

2) **Evaluate the use of RCT's in PFS projects.** RCT's are the gold standard of social science. Only when a program shows through more than one well-constructed RCT that it produces the desired outcomes can it be said without a doubt that there is a causal link between those programs and the outcomes. The inverse is not true; that is, the lack of an RCT does not mean that there is not a causal link. There are other less rigorous evaluation methodologies that are indicative of causation. And tying an RCT to payments in the context of a PFS project adds layers of complexity and timing that may distort the RCT or cause the project to make suboptimal decisions.

3) **Have a clear sense of the quality of available data for both referrals and outcomes.** One of the critical reasons for a pilot is to understand how the data will work in practice. It is not uncommon for administrative data to look accurate and timely when reviewed in a conference room but to turn out to be less instructive when you try to use it in the field. In this project, for example, we could not know that a large percentage of the home addresses listed in Probation's central files (as opposed to the files of individual probation officers) would prove to be inaccurate or dated.

4) **Carefully examine risk assessment tools and the definition of "high risk".** In this project, all people who were referred to treatment or control were "high risk" based on Probation's risk score or the HKS Government Performance Lab score created for the project. These quantitative, records-based estimates of risk were necessary proxies for whether a person was suitable for Roca's intervention, an intervention designed for high-risk young people. It was the only way that the project could run a study that randomized high risk young men to Roca and a control group. On too many occasions, however, the algorithms did not align with Roca's definition of high risk and as a result, complicated Roca's ability to deliver its intervention effectively.

5) **Involve all the government agencies who have a role in operations in planning.** This project was primarily conceived of and managed in the office of the Governor of the Commonwealth and in the Executive Office of Administration and Finance, with the advice of the HKS Government Performance Lab. These were indispensable parties to the ability to launch the project and make the necessary modifications through a decade plus of operations. The agencies that were closer to operations, such as Probation, which in Massachusetts is part of the judicial branch, Corrections, and the separately elected county sheriffs, however, who had valuable insights into data and referrals were not brought to the table early enough.

ATTACHMENT A: INDEPENDENT EVALUATOR FINAL REPORT

Final Report of the Massachusetts Juvenile Justice Pay for Success Evaluation

Akiva Liberman, Ph.D.
Independent Evaluator

May 2024

Abstract

The Massachusetts Juvenile Justice Pay for Success (PFS) project is an experiment in a financing model in which the Commonwealth of Massachusetts would pay the costs of a juvenile justice intervention only after it had been shown to have successfully produced the intended benefits. Initially, other funders – both lenders and grantors – supported the intervention as well as a rigorous independent evaluation. The treatment provider (Roca) also deferred some payments, which were made contingent on the evaluation findings. If the evaluator found the project to have achieved "success" according to contracted metrics initially agreed to by all parties, then the Commonwealth would make "success payments" that would be used to replay lenders with interest and a small return, allow grants to be "redeployed," and make deferred payments to Roca.

The intervention financed by the PFS contract was the use of Roca as a reentry intervention for young men released from incarceration. Roca's four-year intervention consisted of two years of intensive intervention with two years of follow-up. Contracted outcomes of interest, which would generate repayment, were reduced reincarceration and increased employment over the five years following randomization.

The project consisted of a randomized control trial (RCT) in which 1,819 young men were randomized to treatment or control conditions between 2014 and 2018. In addition to estimating the effects of random assignment (RA) to treatment as assigned, the evaluation also estimated the effects of *enrollment* in the program.

The evaluation design also contemplated the possibility that sample sizes assigned or enrollment might fall short of expectations, thereby reducing statistical power – which indeed occurred. Under these circumstances, the evaluation design called for a supplemental quasi-experimental estimate to be conducted, using a difference-in-differences (DID) design.

The separate RCT and DID estimates of the effects of treatment enrollment were then combined (with the RCT and DID estimates weighted at 80% and 20% respectively) into *backstop* estimates, which are the final determinants of payment obligations under the Pay for Success contract.

The final RCT and DID estimates for employment and incarceration are too small relative to their standard errors to provide confidence in either their direction or magnitude. However, the RCT point-estimates were both in the detrimental direction, while the DID point-estimates were in the beneficial direction. When the combined backstop estimates were calculated, neither the recidivism nor employment backstop outcomes were in the beneficial direction, triggering no payment obligations under the PFS contract.

Acknowledgments

This evaluation report is the product of the work of many people, in addition to the many people involved in conceiving, managing, and implementing the project and the Roca intervention.

The evaluation *per se* involved the direct efforts of many people, although I alone bear responsibility for the final product and any errors. The evaluation was designed by Jeffrey Liebman at the Harvard Kennedy School of Government, who also provided feedback along the way. Laura Lempicki of the MA Department of Probation provided critical data and advice on its interpretation. The initial evaluator was Lisa Sonbonmatsu at Sybalitics, and at the Urban Institute, Janine Zweig served as Principal Investigator for the evaluation before I assumed that role. At the Urban Institute, Dan Lawrence, Jennifer Yahner, Jeremy Welsh-Loveman, Rochisha Shulka, and Sara Bastomski were critical in implementing randomization and conducting initial analyses, while Doug Wissoker provided invaluable statistical consultation.

In addition, other project partners provided critical feedback and consultation during the course of the evaluation, notably including John Grossman at Third Sector Capital Partners/YSI, Lili Elkins at Roca, and Mark Attia and Aditya Bashir from MA Administration and Finance.

Table of Contents

Table of Contents

Table of Exhibits

I. OVERVIEW

The Massachusetts Juvenile Justice Pay for Success (PFS) project is an experiment in a financing model in which the Commonwealth of Massachusetts would pay the costs of a juvenile justice intervention only after it had been shown to have successfully produced the intended benefits. Initially, other funders – both lenders and grantors – supported the intervention as well as a rigorous independent evaluation. The treatment provider (Roca) also deferred some payments, which were made contingent on the evaluation findings. If the project were deemed successful according to contracted metrics as assessed by the independent evaluator, then the Commonwealth would make "success payments" that would be used repay the lenders with interest and a small return, allow the grants to be "redeployed," and make deferred payments to Roca.

The intervention financed by the PFS contract was the use of Roca as a reentry intervention for high-risk young men released from incarceration to specified communities in the Boston, Chelsea, or Springfield areas.

Roca' model to help high-risk young men break the cycle of reoffending uses targeted life skills, education, and employment programming, to increase job readiness and workforce participation. Roca's four-year model—which consists of two years of intensive engagement and two years of follow-up—includes four basic elements: relentless outreach to young men by Roca staff; intensive case management; life skills, educational, prevocational and employment programming; and work opportunities with community partners. Roca also connects participants to each other and adults through intensive relationships.

During the course of the project Roca modified its intervention with the aim of improving it and increasing its efficacy, as well as in response to restrictions during the acute phase of the COVID pandemic. In 2016, after a preliminary pilot, Roca restructured its intervention model and grounded it in the delivery of Rewire CBT, [1] to help participants manage chronic and ongoing trauma they have experienced. Because of the timing of this project, 83% of participants were not fully trained and coached in Rewire CBT, and did not receive Roca's intervention model as it is currently delivered.

The beneficial outcomes that would generate repayment under the PFS contract are reduced reincarceration and increased employment over the five years following randomization. Under the PFS contract, success and resulting payments are not all or nothing; rather, the size of payments depend on the number of incarceration days averted and the degree of employment increases. These are assessed in this report.

Basic features of the evaluation design are summarized here; additional detail can be found in the Revised Evaluation Plan, which is included in Appendix F.

---

[1] Rewire CBT – Skills for Living (Rewire) was developed with Massachusetts General Hospital for use in both street and community settings. It was designed to overcome traditional barriers to access, helping front-line staff reach high-risk young people who tend to refuse interventions in clinical settings. Roca's approach involves teaching and modeling of core CBT skills in real time to build: emotional regulation (the ability to manage their emotions and slow down the fight, flight or freeze response), behavior activation (ability to do something different and move out of bottom-brain living), and cognitive flexibility (ability to think differently and make intentional choices.)

RCT Design

From 2014 through 2018, participating agencies referred all young men in the state who seemed to be eligible candidates for the reentry intervention based on criteria of age and high risk.[2] These referrals were transmitted to the independent evaluator (initially Sybalitics, and then the Urban Institute). The evaluator assessed them for eligibility, which was limited to individuals with addresses in specified jurisdictions in the Boston, Chelsea, and Springfield areas. The evaluator then randomized eligible candidate to treatment (Roca) or control conditions and transmitted the names of people randomized to treatment to Roca, who recruited them for participation.

Randomization was stratified within calendar quarters (and months), and the ratio at which individuals were randomized to treatment versus control was varied to accommodate program capacity and the pool of eligible participants, both of which were larger in early quarters. Appendix A shows randomization by calendar quarter.

Estimating the Effects of Random Assignment to Treatment

With this design, a simple comparison of individuals randomized to treatment versus control is inadequate for estimating of effect of treatment, because the randomization ratio varies over time, and that may be confounded with differences over time in outcomes (i.e., criminal behavior, sentencing practices, economic and employment prospects), as well as modifications to the program. The evaluation plan controlled for this confounding of time and randomization through the use of design weights, so that the weighted distribution over time was equivalent for the control and treatment samples.

The regression estimates also included covariates for randomization quarter and other covariates. These estimates of the effects of random *assignment* to treatment are known as Intent to Treat (ITT) estimates.

Estimating the Effect of Enrollment

Participation in the Roca intervention was voluntary. Although Roca's programming is designed for hard-to-engage individuals, it was understood that not all individuals assigned to treatment would participate. Once individuals were randomized to treatment by the evaluator, Roca attempted to recruit them. In addition, Roca was allowed to enroll people who sought out the program themselves, known as "self recruits," regardless of whether they had been randomized to the control condition by the evaluator. (Roca was not given names of people randomly assigned to the control condition.) Instrumental Variable (IV) estimates were used to estimate the effects of enrollment in the program.

Difference in Differences Design and Backstop Estimate

The evaluation design also contemplated the possibility that sample sizes assigned to the control group, or the fraction who enrolled,[3] might fall short of expectations, reducing statistical power – which indeed occurred. Under these circumstances, the evaluation design called for a supplemental quasi-experimental estimate to be conducted using a difference-in-differences (DID) design. The IV estimates from the RCT and DID were then combined into the *backstop*

---

[2] Cases referred from DOP had been prescreened for high risk, based on DOP risk scores, and all Department of Youth Services referrals were high risk. Cases referred from other referral sources were generally not prescreened for risk. Instead, a risk score was computed by the evaluator, and high risk was used as a criterion for eligibility.

[3] More precisely, the criterion depends on the difference in the fraction who enrolled among those assignment to treatment compared to those assigned to control. See discussion of the Effects of Enrollment, on page 3.

estimates, which are the determinants of payment obligations under the PFS contract. The RCT and DID estimates were weighted at 80% and 20% respectively.[4]

Summary of Findings

All of the estimates included 0 in their confidence intervals, so that the evaluation results give no confidence in either the magnitude of direction of effects. The point-estimates from the RCT for both incarceration and employment were in the detrimental direction, that is, more incarceration and less employment. The point-estimates from the DID were in the beneficial direction.

Payment obligations in the PFS contract did not consider standard errors or statistical significance as criteria for payment obligations. Instead, the contracted payment obligations depend on the point estimates of the effects of enrollment.[5] When the RCT's and DID's IV estimates were combined into the backstop estimates, no beneficial effects were found, and no payment obligations were triggered.

II. RANDOMIZED CONTROL TRIAL (RCT) FINDINGS

RCT Methods

Regression was used to estimate the effects of random assignment to treatment on two outcomes, incarceration days and employment. For each outcome, both Intent to Treat (ITT) estimates and Instrumental Variable (IV) estimates are provided. The ITT estimates the effect of being *randomly assigned* to participate in Roca, regardless of treatment participation. The IV estimates focus on the specific effects for individuals who *enrolled* in Roca.

Effects of Enrollment

Generally, the IV estimate controls for failure to enroll among individuals randomly assigned to the treatment condition (as well as controlling for any enrollment among individuals randomly assigned to the control condition). Substantial lack of enrollment among individuals randomized to treatment can diminish the ITT estimate considerably, by including people who receive no actual treatment among the treatment group. By controlling for non-enrollment, the IV estimate more reasonably describe the true effect of the treatment on those individuals who did participate, which are often considerably larger than ITT estimates.

The IV estimate does not adjust for non-participation by trying to model the differences between people who do and do not participate. Rather, conceptually, the IV estimate accommodates the fact that assignment to treatment can only contribute to group differences for those individuals whose participation is affected by whether or not they are assigned to the treatment. Put another way, only those whose who would participate if assigned to the treatment group but would not participate if assigned to the control group can contribute to differences in outcomes between the groups. Thus, the key to adjusting the ITT estimate is based on the *difference* in rates of participation between people assigned to the two conditions. The smaller the difference in participation rates, the fewer individuals actually contribute to the ITT estimates.[6]

---

[4] Revised Evaluation Plan, p. 5, emphasis added: "In the original plan, the RCT was to be weighted at 80% and the DID at 20%. In 2019 Jeffrey Liebman at HKS recommended changing the weights to 70/30, based on new simulations with the final RCT sample sizes. *Now, with the final DID samples in hand, Professor Liebman will be providing updated weights"* – which Dr. Liebman has recommended at 80/20.
[5] Recidivism-reduction also needed to cross a pre-specified threshold in order to generate payments.
[6] These are essentially estimates of the Local Average Treatment Effect, LATE.

All of this means that the actual effect among the people who do participate is diluted by null effects among people who did not participate.[7] However, the IV estimate is no more reliable or precise than the ITT estimate, because the IV estimate's standard error is also considerably larger than the ITT estimate's. In addition, the IV estimate cannot flip the direction of the estimate (except under the very anomalous situation where a *smaller* percentage participate among the group assigned to treatment than among those assigned to control).

RCT Sample

These estimates are based on 1,819 young men who were randomized into the evaluation during the demonstration.[8] Randomization ended in December 2018 (Q21).[9] More people were randomized to Roca in the early quarters (shown in Appendix A), because there was a backlog of eligible participants along with considerable program capacity. Thus, these early quarters contributed more people than later quarters to the evaluation's treatment sample. Per the Evaluation plan, design weights were created for individuals in the control group, so that the ratio of treatment to control group members would be constant over the following variables: time, agency source group type, and geographic area. This balance was checked each quarter. This is described in more detail in the Revised Evaluation Plan in Appendix F.

Of these young men, 1,190 were randomly assigned to the treatment condition, and 447 (37.56%) enrolled in Roca.[10] The other 629 men were assigned to the control condition, and 55 of them (8.74%) enrolled in Roca. The actual enrollment sample involved only 58.4% of the overall Pay for Success contract's aim of enrolling 765 people following randomization.

Cases were referred to the project by the Departments of Probation, Corrections, and Youth Services, as well as state Parole and County Sheriffs. During the RCT, a set of administrative probation cases of individuals in the Boston area was also deemed to be high risk by the project parties after an analysis and thus appropriate for the RCT. The treatment group for the DID, to be discussed later, was limited to the 588 individuals who had been referred from regular probation (i.e., rather than administrative probation) and then randomized to treatment.

[7] The IV estimate here uses enrollment–defined as Roca successfully contacting the individual–as a proxy for participation. But those who enrolled are a superset of those who actually participated, meaning that the ITT estimates are actually even more diluted. Thus, the IV estimate here is conservative in adjusting for participation.

[8] Seven individuals were not included in incarceration outcome results, because they were not matched to any DCJIS records returned when the sample was finalized. In addition, starting with the Q31 results, returned DCJIS records were missing another 6 individuals, having been sealed since the last time their data were obtained. Their most recent outcomes were used.

[9] Another 30 individuals were initially referred to the evaluation and then later deemed ineligible for the study because of prior enrollment in Roca.

[10] An individual was deemed to be enrolled if Roca program staff were successful in contacting him and confirming program eligibility by December 31, 2019, one year after the end of randomization.

Exhibit 1: Sources of Referral to the RCT

	Random Assignment		
Source of Referral to the Evaluation	Control	Treatment	Total
Regular Probation	260	588	848
Administrative Probation	86	120	206
Other State Agencies	142	222	364
County Sheriffs	138	256	394
Total	626	1,186	1,812

Note: This table describes the analytic sample for recidivism.

RCT Observation Period

Outcomes were observed for 5 years following randomization or through March 20, 2023, whichever was sooner. People randomized through June 2017 were observed for a full 5 years, but follow-up time was somewhat shorter for people randomized in the last three quarters of the RCT (Q19, Q20, and Q21). The mean time follow-up time for reincarceration was 1822 days, whereas a full 5 years is 1826 days (365.25 days x 5).

Time to Randomization in the RCT

When the project was designed, it was assumed that the time between reentry from incarceration until randomization would generally be relatively short. Time from reentry to randomization is the sum of (A) time from reentry until referral to the evaluator for possible randomization, plus (B) time from referral to the evaluator until randomization.[11] Indeed, the evaluator was contractually required to randomize individuals to condition within a matter of days after receiving referrals, and this was successfully accomplished.

It was also assumed that the time from reentry until referral to the evaluation (i.e., A) would be relatively short, with the important exception of the initial quarter or two of referrals. At the start of the project, there was a stock of individuals who were already on probation and who were eligible for the intervention. The reentry date was not used as an eligibility criterion, but only that the individuals be on probation (or in the community, if they had been referred from county Sheriffs and were not on probation) at the time of referral and otherwise eligible. Thus, people referred early in the project included both some people who were referred shortly after reentry as well as individuals for whom reentry was longer ago. Of course, once someone was randomized to an RCT condition, they were no longer eligible to be re-randomized. As a result, the project was expected to rapidly “use up” the stock of eligible individuals already on probation in the first project quarter or two, so that in later quarters time from reentry to referral would be short. (The RCT methods included controls for quarter of randomization, as well as design weights, as described earlier, which also prevent differences in time to randomization from biasing RCT results.)

Time to randomization for RCT participants is shown in Exhibit 2. Reentry dates were missing from many individuals who were referred from sources other than the Department of Probation (DOP). In addition, reentry dates were missing from the randomization files for all cases in the first 3 quarters of randomization as well as occasional cases throughout the randomization files.

[11] First, eligibility was determined based on city, age and additional consideration using DOP data as well as using newly requested DJCIS data concerning criminal history. Eligibility criteria are described more fully in the Revised Evaluation Plan in Appendix E.

With the initial quarters excluded, it was anticipated that time to randomization would be fairly short.[12]

Complete data on time to randomization was available for 773 of the 1818 cases in the RCT analyses. Exhibit 3 presents a histogram of time to randomization, which shows a long tail up to about 5 years. As a result, means are considerably longer than medians, and medians are the better summary measure of time to randomization.[13] Exhibit 3 shows that the median time from reentry to referral was 182 days, and from referral to randomization was 15 days.

Exhibit 2: RCT Histogram of Days from Reentry to Randomization

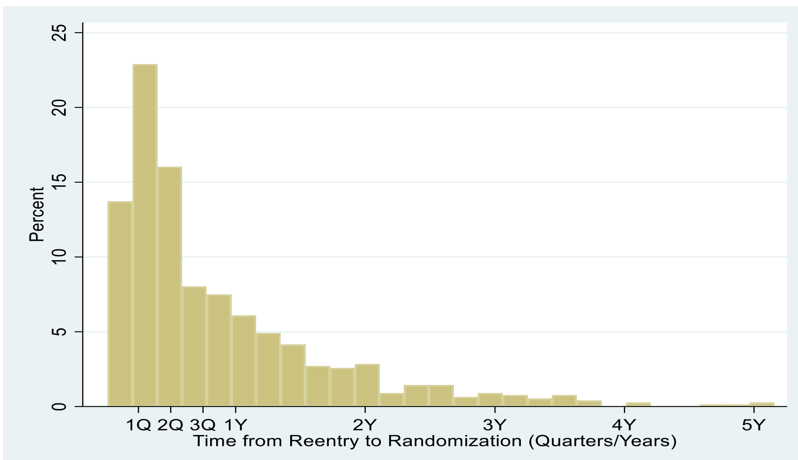


Exhibit 3: RCT Days from Reentry to Randomization

	N	Mean	**Median**	Min	max
Reentry to Referral	773	302.3	**182**	1	1867
Referral to Randomization	773	15.3	**15**	8	31
Reentry to Randomization	773	317.1	**196**	5	1883

[12] Regressing time to randomization – when reentry dates were available -- on assigned condition and randomization quarter (as a continuous variable), and referral agency showed that cases referred by DOP were over 2 weeks longer and cases from the sheriffs were 1 to 4 days quicker than DOP referrals. There was no linear pattern over randomization quarters (with the first 3 quarters all missing).

[13] Time to referral is underestimated slightly, and time from referral is somewhat overestimated, because all referrals are treated as if they were made on the first of every month, whereas in reality the day of the month when referrals were received varied across months and across referral agencies.

Imputation of Incarceration days

For the final RCT estimate, consistent with the evaluation plan, results were imputed in two ways, based on analyses of historical incarceration days before the onset of the project. First, any arraignments that were pending disposition at the end of the observation period had incarceration days imputed for different categories of charged offenses. Second, because the follow-up observation period through March 2023 was less than 5 years for individuals in the last three randomization cohorts (in Qs 19, 20, and 21), their observed incarceration days were increased by preset multiplication factors to allow estimates of a full five-years of follow-up.

COVID-19 and the Evaluation

The current follow-up period includes 2020, 2021, and the first half of 2022, much of which was during the acute phase of the COVID-19 pandemic. COVID-19 affected criminal justice processes in complex ways, especially to reduce transmission of the virus, such as the use of community-based rather than incarcerative sentences, less use of pretrial detention, and/or changes in community supervision and probation requirements. Employment and employment opportunities were also dramatically affected. Simultaneously, criminal behavior changed in complex ways in many U.S. cities, with some reporting declines in property crime, increases in some violent crimes, and surges in drug overdoses; some of these changes persisted even as the most acute phase of COVID-19 passed.

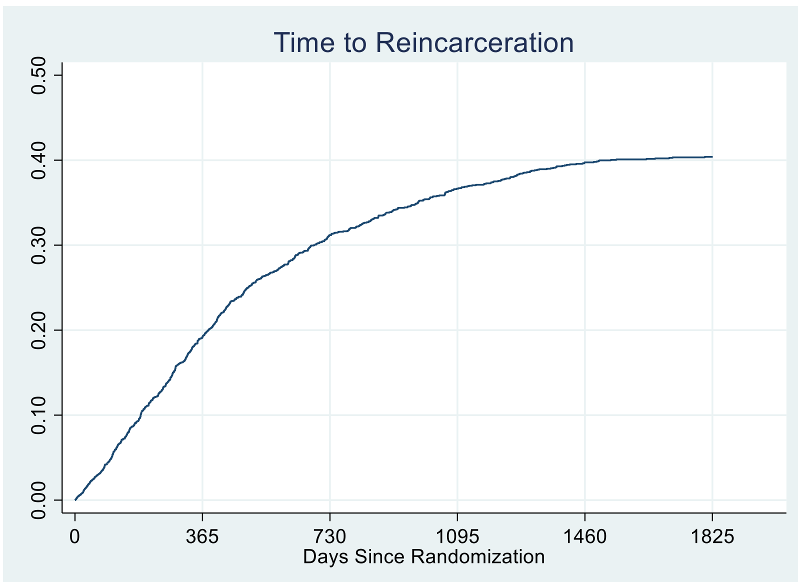
Finally, Roca has reported that program services were delivered during COVID but that they were adjusted from traditional in-person services to virtual services, in order to allow for social distancing. In short, COVID-19 changed the program and likely affected the outcomes being monitored for the evaluation, but we have no *a priori* expectations for how COVID-19 would *differentially* affect outcomes of people in the treatment versus control groups.

RCT Reincarceration Results

Descriptive Results

Time to Arraignments Resulting in Reincarceration
Exhibit 4 shows time from randomization to the first arraignment that resulted in incarceration, for the entire randomization sample. The Y axis shows the cumulative percent who were incarcerated. (This is known as a failure or hazard curve.) By the end of the first year after randomization, the reincarceration rate was not quite 20 percent, rose to about 30 percent by the end of two years of follow-up, and plateaued at around 40 percent by the end of four years.

Exhibit 4. RCT Arraignments Resulting in Reincarceration



Reincarceration by Cohort
Exhibit 5 stratifies the sample by the quarter of randomization. Years are shown in different colors, from blue to red. There may be substantive reasons to expect the first cohort to be somewhat different than later cohorts, and that cohort (randomized in Q2) is highlighted by being the thickest; the last cohort is shown in magenta. On inspection fewer people fail in the first and last cohorts than in other cohorts. More generally, we see considerable variation in the failure rates across cohorts, and the pattern across cohorts seems consistent with those shown in next exhibit, which shows a peak around the cohort randomized in quarter 5.

Exhibit 6 shows the observed incarceration rate at the end of the observation period, broken out by the randomization cohorts. These observed rates of incarceration are the last observed rate – the rightmost point – in the failure curve for that cohort in Exhibit 5.

For individuals with only one observed arraignment after randomization, the incarceration outcome is defined as the longest minimum sentence associated with that arraignment. For individuals with multiple observed arraignments, sentences are summed. To prevent a few people with long sentenced from dominating results, they are capped at a maximum of 10 years per person.

Exhibit 7 shows the mean incarceration days among individuals who were sentenced to incarceration, by randomization cohorts. Exhibit 8 then shows the mean incarceration days for all individuals, including zero days for those not sentenced to incarceration at all.

Exhibit 5. RCT Time to Reincarceration

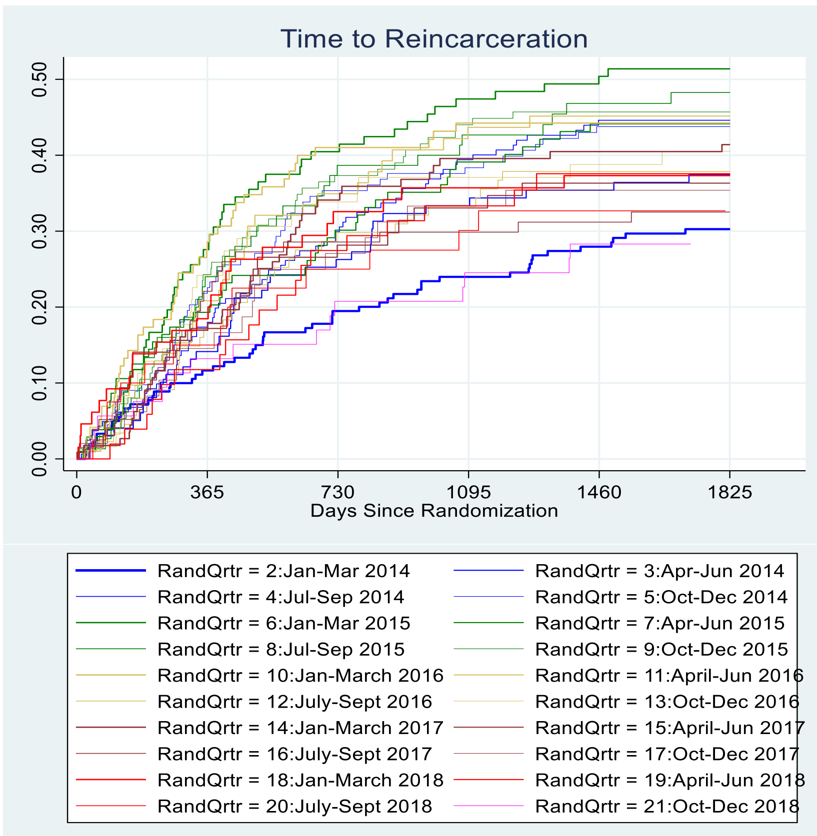


Exhibit 6. RCT Percent with Any Incarceration Days, by Randomization Cohort

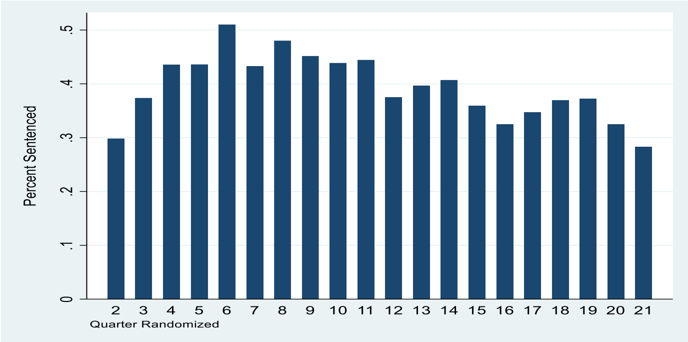


Exhibit 7. RCT Incarceration Days Among Individuals Sentenced to Incarceration, by Quarter Randomized

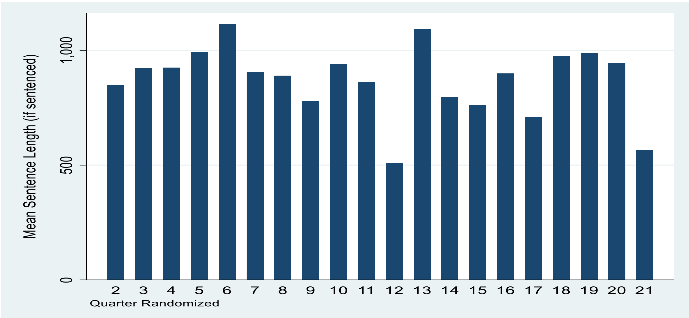
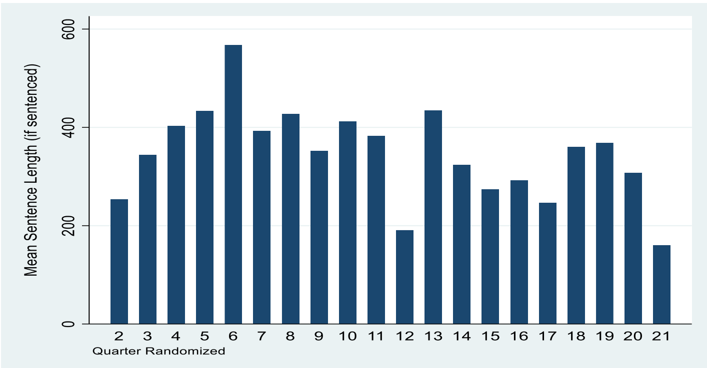


Exhibit 8. RCT Incarceration Days, by Quarter Randomized



Descriptive Results by Condition
We observe that 40% of the RCT sample was sentenced to reincarceration during the follow-up period, including both treatment and control participants. Among those sentenced to reincarceration, the mean sentenced days was 886. Including individuals with no sentences (for whom days equal 0), the mean days of incarceration is 353, with a standard deviation of 676.

Descriptively, without accounting for any control variables and including 0 days, the mean days of incarceration was 329 for the control group and 367 for the treatment group, as shown in Exhibit 9. Among individuals randomly assigned to treatment, 41% were sentenced to incarceration on new charges; if sentenced, their incarceration averaged 897 days. In contrast, among individuals randomly assigned to control, 38% were sentenced to incarceration on new charges; if sentenced, their incarceration day averaged 862 days.

Exhibit 9: RCT Descriptive Results for Incarceration

	Control	Treatment	Total
N	627	1187	1814
N reincarcerated	239	485	724
% reincarcerated	38%	41%	40%
incarceration days if reincarcerated	862	897	886
total incarceration days	329	367	353

These results are uncontrolled and unweighted. Most importantly, the percentages randomized to treatment were intentionally varied over time, with a higher percentage randomized to treatment in the early quarters. This changing randomization rate was confounded with changing recidivism and sentencing rates over time, potentially distorting the simple comparison of individuals randomized to the two conditions. The regression results control this confounding, and estimates are also improved through the inclusion of a set of covariates.

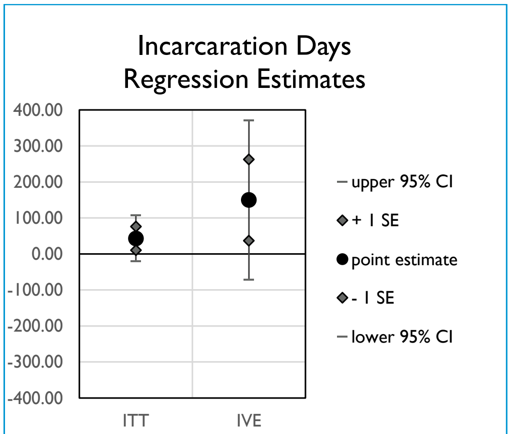
Regression Results for Incarceration Days
The outcome estimated is incarceration days resulting from arraignments on new charges since randomization. Because the regressions include statistical controls for quarters of randomization, in effect comparisons are made between treatment and control individuals who were randomized during the same quarter.

Intent to Treat Estimate of the Effect of Random Assignment to Roca
The ITT estimate is that random assignment to Roca was associated with an increase of 43 days of incarceration, with a standard error of 33 days. The 95-percent confidence interval (±1.96 standard errors) ranges from 21 fewer days to 108 additional days of incarceration.

Instrumental Variable Estimate of the Effect of Enrollment in Roca
Compared to the ITT estimate, the IV estimate increases the size of the estimated effect and its standard error. We find that enrollment in Roca was associated with an increase in 151 days of incarceration, with a standard error of 113 days. The 95-percent confidence interval ranges from 70 fewer days to 372 additional days of incarceration.

Regression output is shown in Appendix B, Exhibit 22 and Exhibit 23.

Exhibit 10. RCT Illustration of Incarceration Days Regression Estimates



Logistic Regression Estimate of Assignment on the Odds of Incarceration
An additional ITT estimate analysis was also conducted as a sort of sensitivity test on these incarceration results. To eliminate the possibility that the preceding results are driven by cases with atypically long incarcerations, a logistic regression model was run to estimate the *odds* of incarceration (using the same covariates and weights as in the preceding ITT estimate model). In this model, each individual's outcome is binary (reincarcerated or not), and results estimate how assignment to Roca affects the odds of incarceration. Because this model is not sensitive to long versus short sentences, it does not produce results in the metric of days that is needed to estimate contracted payments. However, it does eliminate any undue influence of a few long sentences and provides a sensitivity test on the *direction* of the previous incarceration results.[14]

The logistic regression model is included in Appendix B, Exhibit 24. Results are shown as odds ratios (ORs). An OR of 1.0 means no effect, ORs greater than 1.0 indicate an increase in incarceration, and ORs less than 1.0 indicate a decrease in incarceration. Results are consistent with the preceding models both in finding that assignment to Roca is associated with slightly higher odds of incarceration (OR = 1.16), and that the possibility of no effect (i.e., OR = 1.0) is within its 95% confidence interval (with ORs ranging from 0.92 to 1.47).[15]

---

[14] The problem of a few long sentences was anticipated in the evaluation design, which capped days sentenced at 10 years, so that all sentence outcomes longer than 10 years are treated as 10 years. 23 RCT participants accrued sentenced days of at least 10 years, and about 60 percent were among individuals randomized to treatment.
[15] ORs have a complex relationship to changes in probability; as an illustration, this OR would raise a 0.300 probability of incarceration to 0.333. For more on interpreting ORs in logistic regression, see Liberman, A. M. (2005). How much more likely? The implications of odds ratios for probabilities. *American Journal of Evaluation*, *26,* 253-266.

RCT Employment Results

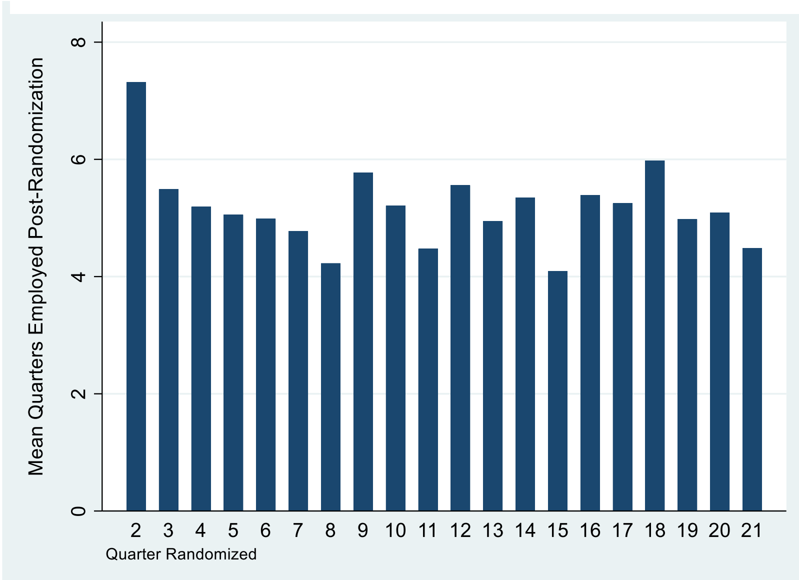
The second outcome analyzed is the number of quarters that an individual was employed and reported income of at least $1,000. Employment outcome data was provided to the evaluator averaged across small groups of 6-12 individuals,[16] rather than for each individual in the sample. That process resulted in 278 groups. These grouped employment outcomes were then attributed back to the individuals in the groups. Regressions to examine the impact on employment were run at the group level.[17]

Descriptive Results

The mean number of employed quarters after randomization was 5.34, with a standard deviation of 2.42.[18] among the entire RCT sample including both treatment and control participants. Descriptively, without accounting for any control variables, the control group's mean observed quarters of employment is 5.35 and the treatment group's mean is 5.32. Exhibit 11 displays the average quarters of employment for the different randomization cohorts. Notably, individuals randomized during the earliest quarter (aka Q2) had longer employment during the following five years than other randomization cohorts.

[16] Groups were comprised of individuals who were identical in both random assignment condition and enrollment status. Beginning with the Q27 estimate, groups were also constrained to be identical on referral source (DOP vs. others). In addition, individuals were grouped to be as similar as possible on quarter of randomization, geographic service area, and age.
[17] 127 individuals were lacking complete social security numbers (SSN) and therefore employment data could not be retrieved for them. These individuals were assigned to groups after the data request and then given the same group mean employment values as those in the group who had SSNs.
[18] These means are at the individual level.

Exhibit 11. RCT Average Quarters of Employment, by Quarter Randomized



Regression Results for Employment

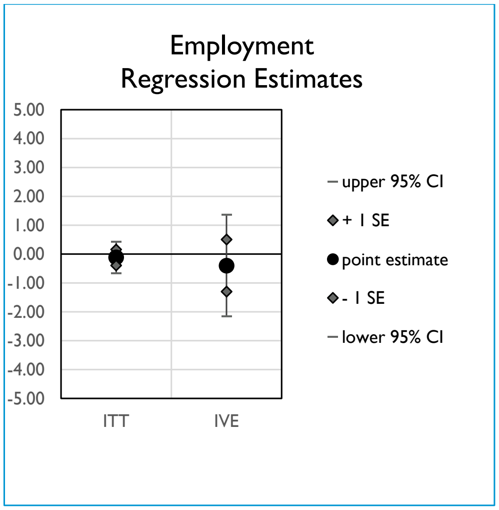
Intent to Treat Estimate of the Effect of Random Assignment to Roca

The ITT analysis estimates that random assignment to treatment decreased employment by 0.12 quarters, with a standard error of 0.28. This generates a wide range of possible effects, in that the 95 percent confidence interval ranges between 0.66 fewer quarters of employment to 0.43 additional quarters of employment.

Instrumental Variable Estimate of the Effect of Enrollment in Roca

The IV analysis estimates that Roca enrollment decreased employment by 0.40 quarters, with a standard error of 0.90. The 95 percent confidence interval ranges from 2.16 fewer quarters of employment to 1.37 additional quarters. This wide range of possible effects provides little evidence of either the size or the direction of the effect of Roca on employment. These results are illustrated in the next Exhibit. Regression output is shown in Appendix B, Exhibit 25 and Exhibit 26.

Exhibit 12. RCT Regression Results for Employment



RCT Estimates Over the Course of the Project

The effects of random assignment (ITT estimate) and enrollment (IV estimate) were estimated 13 times during the course of the project as additional individuals entered the randomized study and follow-up time for individuals in the study increased. Appendix E displays how the estimates changed over time. Incarceration results bounced around a bit in the early estimates and then stabilized, while employment results were remarkably steady.

III. DIFFERENCE IN DIFFERENCES (DID) ANALYSES

Difference in Differences Methods

For the DID, three quasi-experimental comparison groups were assembled from comparison cities (Worchester, Lawrence, Fall River, Brockton) and prior cohorts (reentry 2010-2011). Combined with the treatment group, the four DID groups are organized as shown in Exhibit 13:

Exhibit 13: Difference in Differences Groups

	A. Roca Cities individuals returning to Boston, Chelsea, Springfield	**B Comparison Cities** individuals returning to comparison jurisdictions: Worchester, Lawrence, Fall River, Brockton
I. RCT Period (aka Contemporary Period) Referrals to RCT 2014-2018	**I.A** **Treatment Group**	**I.B** **Contemporaneous comparison group** (in comparison cities) *group E in the Revised Evaluation Plan, p.3:*
II. Historical Period prior cohorts of young men whose reentry preceded the PFS program, with reentry (i.e., probation start dates) in 2010 and 2011[19]	**II.A** **Historical comparison group** (in the Roca cities): *group F in the Revised Evaluation Plan*	**II.B** **"Diagonal Comparison Group"** i.e. Historical comparison group in the comparison cities**:** *group G in the Revised Evaluation Plan*

Because referrals of candidates for the RCT were made statewide, and the evaluator then identified individuals in the Boston, Chelsea, and Springfield areas for the RCT, these same referral files were used to identify individuals who were candidates for the contemporaneous comparison group from comparison cities. That is, the contemporary comparison group had been referred prospectively for the evaluation. In contrast, the cohorts of young men in the historical comparison groups were identified retrospectively by DOP for the evaluation.

To render the groups comparable, the treatment group for the DID design was also limited in various ways to a subset of the RCT treatment group. Most importantly, because DOP identified the individuals for the historical comparison groups for the evaluation, the DID's treatment group was also limited to individuals who had been referred to the RCT by the DOP (see Exhibit 1).

---

[19] RCT participants (i.e., the treatment group) were referred to the evaluation from 2014 through 2018. The Revised Evaluation Plan assumed that their probation start dates would reach back into 2013. The historical comparison groups were to be composed of young men with probation start dates in the years preceding the treatment group, namely from 2010 through 2012. However, on inspection, a considerable number of the RCT participants had probation start dates in 2012, which is consistent with the findings concerning time to randomization already shown in Exhibit 3. As a result, young men with probation start dates in 2012 were excluded from the DID comparison cohorts (i.e., groups II.A and II.B above), which was instead restricted to 2010 and 2011.

To estimate the effect of treatment, the DID examines the interaction between Design Period and Design Jurisdiction. That is, it examines how the difference in outcomes between jurisdictions itself varies by the two time periods. Put another way, the outcomes are compared as follows:

treatment effect = (I.A – I.B) – (II.A– II.B)

Details of this design are further described in the Revised Evaluation Plan in Appendix F.

Propensity Score Weighting

In quasi-experimental designs, random assignment is not the basis for membership in the treatment versus comparison groups. Thus, even though we try to create groups as similar as possible to the treatment group, we have no guarantee that the groups are equivalent. Yet, unbiased estimates require that the groups be equivalent, posing a conundrum. Propensity scores were developed to address this issue. The propensity score for a case is the probability that the case is a treatment case, with the prediction based solely on preexisting information. Higher scores indicate a greater probability that the case is in the treatment group. The more similar the two groups are, the harder it is to predict someone's group membership from preexisting information.

Propensity scores are used in two ways in the present study. First, when individuals' propensities – i.e., predicted probabilities of being in one group or the other -- depart too much from 50/50, they are "trimmed" (i.e., excluded). (These cases are referred to as being "off common support.") The evaluation plan thresholds for trimming an individual as having odds too high of the treatment were that odds exceeded 5.0 and/or exceeded all individuals in the comparison group); the threshold for odds being too low were that odds were under 0.20[20] and/or below all individuals in the treatment group.[21]

Second, "balanced" groups have equivalent distributions on the propensity scores. Propensity scores are used to weight the comparison group so that its distribution on the propensity score is similar to the distribution found in the treatment group.

In our DID design, each comparison group had its own propensity score comparing it to the treatment group. That propensity score was used to trim individuals from that group and from the treatment group. After trimming, the remaining cases were weighted. Propensity scores were developed using logit models, which are included in Appendix C.[22]

---

[20] These odds correspond to probabilities (*p*s*)* below 16.6% or above 83.3%. Odds = p/(1-p); p= odds/(odds+1). People were excluded if their *propensity weights* were below 0.20 or above 5.0. The propensity weights here (ATT weights) are the odds of being in the treatment group.

[21] These odds thresholds are appropriately symmetric when the treatment and comparison groups are the same size. (When one group is appreciably larger than the other, the odds of being in the larger groups increase.) Because group sizes were unequal, the comparison groups were upweighted to be the same size as the treatment group.

[22] Contrary to the Revised Evaluation Plan, the propensity score model for the contemporaneous comparison group (i.e., group I.B above) did not include the quarter in which individuals were referred to the evaluation (a.k.a. randomization quarter). The plan correctly notes that including dummy variables for randomization quarter would accomplish what the RCT accomplished by using design weights, namely, produce equivalent distributions over time. However, including the randomization quarters in the PS would also lead to excluding individuals from the DID, particularly from quarters that were unusual in the number of cases randomized to treatment (especially the early quarters) – which was not the intention. Instead, as in the RCT, design weights were created and applied to the contemporaneous comparison group to equalize its distribution over calendar quarters to the distribution of DID treatment cases.

The number of people trimmed from each group and the resulting sample sizes are shown in Exhibit 14, where the rightmost group is the treatment group. The propensity scores for the jurisdictional comparison only resulted in trimming 9 people (3 and 6, respectively, from the comparison and treatment groups). The propensity scores for the two historical comparison groups required trimming of more individuals. Ultimately 501 individuals from the treatment group were retained for analyses, while the comparison groups consisted of 654 to 1,071 individuals.

Exhibit 14: DID Sample Sizes following Propensity Score Weighting

	Historical Period		Contemporary Period		Total
Propensity Score Trimming	Comparison Cities	Roca Cities	Comparison Cities	Roca Cities	
ok	**654**	**1,071**	**838**	**501**	**3,064**
Jurisdiction comparison			3	6	9
Historical comparison		45		53	98
Diagonal comparison	33			17	50
Total before trimming	687	1,16	841	577	

Note: People in the treatment group who had more than one reason to be trimmed are shown by the first reason for trimming in the order shown from top down.

Enrollment Among the Treatment Group

Among the 501 individuals retained in the treatment group, 192 enrolled in Roca (38.3%). The DID's ITT estimates are divided by this proportion to produce the DID's IV estimate, in effect multiplying the point estimates by 2.61.

Time from Reentry to Randomization for the DID treatment Group

Because the DID treatment group is a subset of the RCT treatment group limited to DOP referrals, time to randomization was reexamined for the DID treatment group with reentry dates. Median time to randomization was a few days shorter than for the RCT overall, shown earlier.

Exhibit 15: DID Days from Reentry to Randomization

	N	Mean	**Median**	Maximum
Reentry to Referral	265	267	**170**	1352
Referral to Randomization	265	15	**15**	24
Reentry to Randomization	265	282	**186**	1363

To make the historical groups' follow-up periods comparable to the contemporary groups, the follow up periods for the historical cohorts were begun 186 days after reentry, the median number of days from reentry to randomization.

Design Modifications

In operationalizing the DID design and estimation, several modifications were made to what was described in the Evaluation plan. These modifications were described in footnotes 19, 21 and 22.

DID Results

Descriptive Results

Exhibit 16 displays time to the first sentence of incarceration for the DID sample (after trimming); the Y-axis shows the cumulative percent who had been reincarcerated. As can be seen, after one year of follow-up, over 30% had been reincarcerated, and by five years that had risen to above 40%.

Exhibit 16: DID Time to Reincarceration

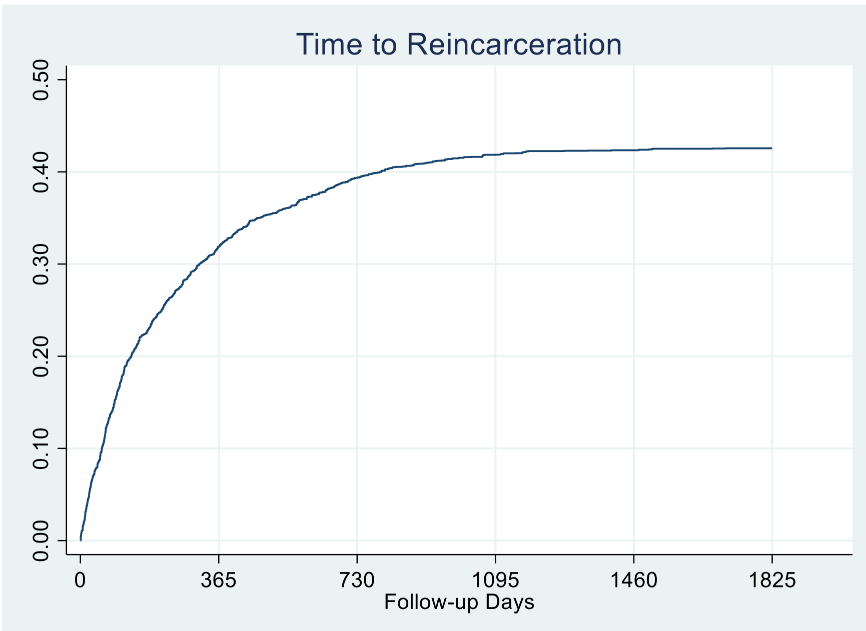


Exhibit 17 displays descriptive results for the groups, after trimming; the treatment group is rightmost. These results are not weighted by propensity scores or design weights, nor controlled for any covariates. The observed percent reincarcerated ranged from 36% to 43%, and were lower in the Roca-serving cities than in the comparison cities, in both time periods. If reincarcerated, individuals received total sentences averaging 760 days, with groups varying between 711 to 838 days. The treatment group, which is the rightmost, had at the same time the *smallest* fraction of individuals reincarcerated as well as the *longest* total sentences if incarcerated. Group averages for sentenced days, including 0s for those not reincarcerated, ranged from 301 to 330 days, with the treatment group coming in lowest.

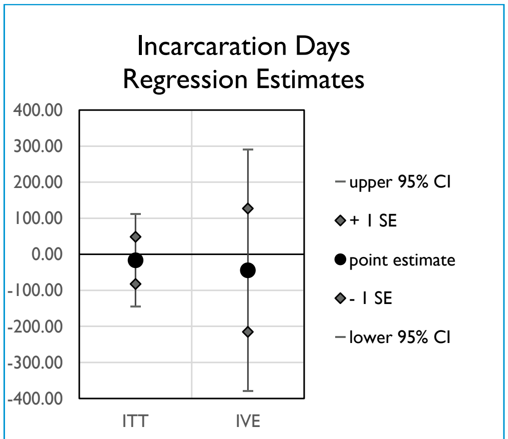
Exhibit 17: DID Descriptive Results for Incarceration

	Historical Period		Contemporary Period		Total
	Comparison Cities	Roca Cities	Comparison Cities	Roca Cities	
N	654	1071	838	501	3064
N reincarcerated	304	459	356	180	1299
% reincarcerated	46%	43%	42%	36%	42%
incarceration days if reincarcerated	711	788	727	838	760
total incarceration days	330	338	309	301	322

Regression Results

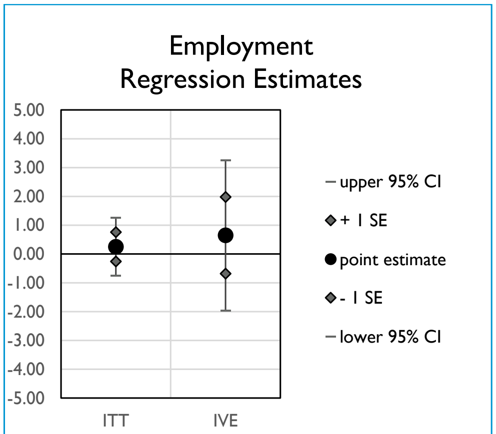
Regression models for the DID are included in Appendix D. For incarceration, the ITT analyses estimated that assignment to treatment decreased incarceration days by 17 days, with a standard error of 65 days. The IV estimate divides estimates by 0.38, thereby magnifying it to a decrease of 44 days. As can be seen in Exhibit 18, these results are quite close to zero, but in the beneficial direction.

Exhibit 18: DID Regression Results for Incarceration



Employment results are displayed in Exhibit 19.The ITT analysis estimated that assignment to treatment increased employment by 0.25 quarters, with a standard error of 0.512. The IV analysis then estimated that enrollment increased employment by 0.65 additional quarters. However, these estimates, too, are well within confidence intervals.

Exhibit 19: DID Regression Results for Employment



IV. BACKSTOP ESTIMATE AND CONCLUSION

The RCT estimates for both recidivism and employment were in the detrimental direction, while both DID estimates were in the beneficial direction – although all estimates included 0 in their confidence intervals.

Exhibit 20: Final Project Backstop Estimates

IV estimate	Incarceration	Employment
RCT (80%)	+151	-0.40
DID (20%)	-44	+0.65
backstop	+112	-0.19

Exhibit 20 shows the RCT and DID combined into the backstop estimate, weighted 80/20 respectively (see footnote 4). The combined backstop estimates are that Roca resulted in 112 days of additional incarceration, and 0.19 fewer quarters of employment. Payment obligations in the PFS contract did not consider standard errors or statistical significance as criteria for payment obligations. Instead, the contracted payment obligations depend merely on the point estimates of the effect of enrollment. Because neither backstop estimate was beneficial, no payment obligations were triggered.

APPENDICES

APPENDIX A: RANDOM ASSIGNMENT

Random Assignment

In particular, a higher fraction of referrals was randomized to the treatment group in early quarters because treatment slots were most available at the beginning of the project, as shown in Exhibit 21. This means that on average the treatment group has longer follow-up observation periods, which leads to greater opportunity to accrue both sentenced days of incarceration and more quarters of employment.

The regression models control for this bias in two complementary ways. First, the models include terms controlling for the quarter of randomization, so that model results reflect a comparison of people randomized to the two conditions in the same quarter, which is then pooled over all quarters. Second, weights are applied to control participants so that their weighted distribution over time resembles that of treatment participants.

Exhibit 21. RCT Analytic Sample by Quarter and Treatment Status

	Control	Treatment	Total
Q2: Jan-Mar 2014	39	142	181
Q3: Apr-Jun 2014	23	77	100
Q4: Jul-Sep 2014	41	122	163
Q5: Oct-Dec 2014	35	99	134
Q6: Jan-Mar 2015	24	78	102
Q7: Apr-Jun 2015	28	76	104
Q8: Jul-Sep 2015	18	57	75
Q9: Oct-Dec 2015	35	90	125
Q10: Jan-Mar 2016	45	53	98
Q11: Apr-Jun 2016	30	42	72
Q12: Jul-Sep 2016	33	55	88
Q13: Oct-Dec 2016	29	34	63
Q14: Jan-Mar 2017	52	61	113
Q15: Apr-Jun 2017	29	35	64
Q16: Jul-Sep 2017	38	39	77
Q17: Oct-Dec 2017	23	26	49
Q18: Jan-Mar 2018	34	32	66
Q19: Apr-Jun 2018	25	27	52
Q20: Jul-Sep 2018	20	20	40
Q21: Oct-Dec 2018	28	25	53
Total	629	1,190	1,819

APPENDIX B: RCT REGRESSION OUTPUT

RCT Incarceration Days

The regression results for days of incarceration are shown in Exhibit 22 and Exhibit 23.[23] The regression models are least squares models, so that the coefficients are readily interpretable as the number of additional days of incarceration associated with treatment.

Exhibit 22 shows the Intent to Treat (ITT) estimate, which compares the outcome for people who were randomized to treatment versus randomized to control.

Exhibit 22. RCT ITT estimate Estimate for Incarceration Days (RecidDays)

```
Linear regression                                   Number of obs     =      1,812
                                                    F(31, 1780)       =       5.82
                                                    Prob > F          =     0.0000
                                                    R-squared         =     0.1319
                                                    Root MSE          =     628.51
```

RecidDaysQ40	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
RandQrtr						
3:Apr-Jun 2014	-9.717315	74.86834	-0.13	0.897	-156.5564	137.1218
4:Jul-Sep 2014	-29.75276	73.95214	-0.40	0.687	-174.7949	115.2894
5:Oct-Dec 2014	-1.309312	78.01363	-0.02	0.987	-154.3173	151.6986
6:Jan-Mar 2015	183.36	103.1484	1.78	0.076	-18.94474	385.6648
7:Apr-Jun 2015	36.95799	76.92892	0.48	0.631	-113.9225	187.8385
8:Jul-Sep 2015	64.07834	102.3432	0.63	0.531	-136.6471	264.8038
9:Oct-Dec 2015	16.48769	70.29021	0.23	0.815	-121.3723	154.3477
10:Jan-March 2016	92.31066	76.03142	1.21	0.225	-56.80959	241.4309
11:April-Jun 2016	6.699813	86.10926	0.08	0.938	-162.1861	175.5857
12:July-Sept 2016	-163.431	56.21553	-2.91	0.004	-273.6864	-53.17563
13:Oct-Dec 2016	52.82812	96.63994	0.55	0.585	-136.7116	242.3678
14:Jan-March 2017	-46.17847	71.34169	-0.65	0.518	-186.1008	93.74381
15:April-Jun 2017	-62.15887	71.68546	-0.87	0.386	-202.7554	78.43765
16:July-Sept 2017	-2.013955	91.94436	-0.02	0.983	-182.3442	178.3163
17:Oct-Dec 2017	-62.13754	77.43192	-0.80	0.422	-214.0046	89.7295
18:Jan-March 2018	46.81445	104.0058	0.45	0.653	-157.1719	250.8008
19:April-Jun 2018	-1.092521	132.8277	-0.01	0.993	-261.6072	259.4221
20:July-Sept 2018	-28.03089	111.7375	-0.25	0.802	-247.1815	191.1197
21:Oct-Dec 2018	-147.2708	67.91488	-2.17	0.030	-280.472	-14.06945
Agency						
2:Referred by nonprobation	-9.038115	58.88885	-0.15	0.878	-124.5367	106.4605
City						
2:Chelsea Roca Area	53.33185	47.34513	1.13	0.260	-39.52605	146.1897
3:Springfield Roca Area	-39.10945	36.75597	-1.06	0.287	-111.1989	32.97996
Agency#City						
2:Referred by nonprobation#2:Chelsea Roca Area	65.635	98.5417	0.67	0.505	-127.6346	258.9046
2:Referred by nonprobation#3:Springfield Roca Area	160.4023	88.98578	1.80	0.072	-14.12533	334.9299
PFSAge	-61.65747	10.8073	-5.71	0.000	-82.8538	-40.46114
Age1st_arr	-14.3988	7.446255	-1.93	0.053	-29.00313	.2055197
Arr_pre_tot_PCF	11.32593	5.00567	2.26	0.024	1.508321	21.14354
Inc_pre_tot_PCF	2.200192	11.2982	0.19	0.846	-19.95894	24.35932
dataQ40_persTotFDISP_pre	.2284617	.047254	4.83	0.000	.1357824	.321141
meanPreEmpl	-25.07628	17.07637	-1.47	0.142	-58.56812	8.415556
TxAssigned	43.44689	32.68072	1.33	0.184	-20.64973	107.5435
_cons	1691.594	274.1277	6.17	0.000	1153.948	2229.24

Note: The key term in the model is TxAssigned.

[23] Criminal history records could not be obtained for seven individuals in the sample.

Exhibit 23 shows the Instrumental Variable Estimate (IV estimate), which estimates the outcome for individuals who are randomized to treatment and enroll in Roca, as compared to comparable individuals in the control group who are randomized to control and do not enroll in Roca. (This is the second stage of a 2SLS regression.)

Exhibit 23. RCT IV estimate for Incarceration Days (RecidDays)

Wald chi2(31)	=	184.51
Prob > chi2	=	0.0000
R-squared	=	0.1254
Root MSE	=	625.27

RecidDaysQ40	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
TxEnrolled	150.6565	112.8161	1.34	0.182	-70.45906	371.772
RandQrtr						
3:Apr-Jun 2014	-12.88635	75.56607	-0.17	0.865	-160.9931	135.2204
4:Jul-Sep 2014	-34.86315	73.87945	-0.47	0.637	-179.6642	109.9379
5:Oct-Dec 2014	13.75779	77.46314	0.18	0.859	-138.0672	165.5828
6:Jan-Mar 2015	196.4116	101.552	1.93	0.053	-2.626761	395.4499
7:Apr-Jun 2015	31.22492	74.8757	0.42	0.677	-115.5288	177.9786
8:Jul-Sep 2015	64.37678	101.159	0.64	0.525	-133.8912	262.6447
9:Oct-Dec 2015	25.71004	69.93241	0.37	0.713	-111.355	162.775
10:Jan-March 2016	94.14551	75.26017	1.25	0.211	-53.36171	241.6527
11:April-Jun 2016	3.987007	84.65325	0.05	0.962	-161.9303	169.9043
12:July-Sept 2016	-164.3231	56.15325	-2.93	0.003	-274.3815	-54.26479
13:Oct-Dec 2016	45.76136	96.25452	0.48	0.634	-142.894	234.4168
14:Jan-March 2017	-46.98048	71.16458	-0.66	0.509	-186.4605	92.49954
15:April-Jun 2017	-57.86464	71.72442	-0.81	0.420	-198.4419	82.71265
16:July-Sept 2017	11.99488	90.83373	0.13	0.895	-166.036	190.0257
17:Oct-Dec 2017	-58.65671	77.83161	-0.75	0.451	-211.2039	93.89044
18:Jan-March 2018	73.84431	106.9848	0.69	0.490	-135.8421	283.5307
19:April-Jun 2018	4.031657	132.4544	0.03	0.976	-255.5743	263.6376
20:July-Sept 2018	-16.8613	110.9284	-0.15	0.879	-234.277	200.5544
21:Oct-Dec 2018	-119.6493	72.37183	-1.65	0.098	-261.4955	22.19684
Agency						
2:Referred by nonprobation	-6.583832	58.73213	-0.11	0.911	-121.6967	108.529
City						
2:Chelsea Roca Area	52.94005	47.27946	1.12	0.263	-39.72598	145.6061
3:Springfield Roca Area	-40.00776	36.69866	-1.09	0.276	-111.9358	31.9203
Agency#City						
2:Referred by nonprobation#2:Chelsea Roca Area	64.30246	97.70287	0.66	0.510	-127.1917	255.7966
2:Referred by nonprobation#3:Springfield Roca Area	165.7461	88.00052	1.88	0.060	-6.73172	338.224
PFSAge	-58.60459	11.50282	-5.09	0.000	-81.14971	-36.05947
Age1st_arr	-14.24284	7.40426	-1.92	0.054	-28.75492	.2692414
Arr_pre_tot_PCF	12.0574	4.996414	2.41	0.016	2.264603	21.85019
Inc_pre_tot_PCF	1.897914	11.13441	0.17	0.865	-19.92514	23.72097
dataQ40_persTotFDISP_pre	.2248343	.0475125	4.73	0.000	.1317115	.317957
meanPreEmpl	-13.14626	20.27214	-0.65	0.517	-52.87892	26.58641
_cons	1584.84	307.9811	5.15	0.000	981.2085	2188.472

Instrumented: TxEnrolled

Note: The key term in the model is TxEnrolled.

Exhibit 24. RCT ITT estimate Estimate on the Odds of Incarceration (AnyRecidDays)

```
Logistic regression                                     Number of obs =  1,812
                                                        Wald chi2(31) = 185.87
                                                        Prob > chi2   = 0.0000
Log pseudolikelihood = -1394.7533                       Pseudo R2     = 0.1246
```

anyDaysQ40	Odds ratio	Robust std. err.	z	P>|z|	[95% conf.	interval]
RandQrtr						
3:Apr-Jun 2014	1.188505	.4448642	0.46	0.645	.5706787	2.475202
4:Jul-Sep 2014	.8844199	.2732237	-0.40	0.691	.4827218	1.620392
5:Oct-Dec 2014	.9785148	.3163382	-0.07	0.946	.5192632	1.843942
6:Jan-Mar 2015	1.507563	.4933977	1.25	0.210	.7937666	2.863242
7:Apr-Jun 2015	1.34988	.4665668	0.87	0.385	.6856296	2.657669
8:Jul-Sep 2015	1.345466	.4605338	0.87	0.386	.6878896	2.631643
9:Oct-Dec 2015	1.36177	.4248756	0.99	0.322	.7387997	2.51004
10:Jan-March 2016	1.508089	.4983485	1.24	0.214	.7891272	2.882085
11:April-Jun 2016	1.195306	.427024	0.50	0.618	.5934535	2.407529
12:July-Sept 2016	.9311182	.3097167	-0.21	0.830	.485144	1.787059
13:Oct-Dec 2016	.9794578	.3541083	-0.06	0.954	.4822215	1.989413
14:Jan-March 2017	1.177915	.3695095	0.52	0.602	.6369333	2.178383
15:April-Jun 2017	.9444553	.3376161	-0.16	0.873	.4687064	1.903101
16:July-Sept 2017	.9335237	.3296449	-0.19	0.846	.4672507	1.865094
17:Oct-Dec 2017	1.088312	.4396386	0.21	0.834	.493059	2.402194
18:Jan-March 2018	1.080159	.3962524	0.21	0.834	.5262927	2.216909
19:April-Jun 2018	.8717729	.3541245	-0.34	0.735	.3932197	1.932731
20:July-Sept 2018	.7386584	.3313187	-0.68	0.499	.3066513	1.779273
21:Oct-Dec 2018	.6876289	.2941798	-0.88	0.381	.2973001	1.590425
Agency						
2:Referred by nonprobation	.8923774	.1785619	-0.57	0.569	.6028729	1.320904
City						
2:Chelsea Roca Area	1.228303	.2660605	0.95	0.342	.8033931	1.877944
3:Springfield Roca Area	1.215771	.2454409	0.97	0.333	.8184847	1.805897
Agency#City						
2:Referred by nonprobation#2:Chelsea Roca Area	.8702775	.2705464	-0.45	0.655	.4731966	1.600567
2:Referred by nonprobation#3:Springfield Roca Area	1.047901	.2997346	0.16	0.870	.5982022	1.83566
PFSAge	.7523225	.0285127	-7.51	0.000	.6984638	.8103343
Age1st_arr	.9978011	.0302203	-0.07	0.942	.9402942	1.058825
Arr_pre_tot_PCF	1.067503	.0206545	3.38	0.001	1.027779	1.108762
Inc_pre_tot_PCF	1.156259	.0476638	3.52	0.000	1.066514	1.253556
dataQ40_persTotFDISP_pre	1.000627	.0001099	5.70	0.000	1.000411	1.000842
meanPreEmpl	.8957668	.069849	-1.41	0.158	.7688136	1.043684
TxAssigned	1.162568	.1403413	1.25	0.212	.9176228	1.472898
_cons	89.52161	72.86059	5.52	0.000	18.16122	441.2766

Note: **_cons** estimates baseline odds.

Note: The key term is TxAssigned.

RCT Employment Quarters

The regression results for quarters of employment are shown in Exhibit 25 and Exhibit 26. The regression models are the same as those used for the incarceration outcome. Exhibit 25 shows the ITT estimate, and Exhibit 26 shows the IV estimate, from the second stage of a 2SLS regression.

Exhibit 25. RCT ITT estimate Estimate for Employment (meanPostEmpl)

Linear regression

Number of obs	=	278
F(31, 246)	=	5.36
Prob > F	=	0.0000
R-squared	=	0.4097
Root MSE	=	2.0167

meanPostEmpQ40	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
meanQ2	1.747608	.8052221	2.17	0.031	.1615992	3.333617
meanQ3	2.142624	1.097837	1.95	0.052	-.0197357	4.304984
meanQ4	1.672904	.8900093	1.88	0.061	-.0801068	3.425914
meanQ5	1.049129	.7599432	1.38	0.169	-.4476958	2.545955
meanQ6	1.947935	.7319365	2.66	0.008	.5062732	3.389597
meanQ7	1.226092	.8055616	1.52	0.129	-.3605858	2.81277
meanQ8	1.277347	1.059835	1.21	0.229	-.8101606	3.364855
meanQ9	1.909845	.9115366	2.10	0.037	.1144328	3.705256
meanQ10	1.368891	.8680585	1.58	0.116	-.3408839	3.078666
meanQ11	1.261986	1.025721	1.23	0.220	-.7583292	3.282302
meanQ12	1.871924	.7427655	2.52	0.012	.4089325	3.334915
meanQ13	1.570308	.9527154	1.65	0.101	-.3062123	3.446828
meanQ14	1.556818	.8205136	1.90	0.059	-.0593101	3.172946
meanQ15	-.5636331	.9661796	-0.58	0.560	-2.466673	1.339407
meanQ16	1.175226	.8909264	1.32	0.188	-.5795905	2.930043
meanQ17	.7800599	.9284753	0.84	0.402	-1.048715	2.608835
meanQ18	2.228125	.9568909	2.33	0.021	.3433814	4.11287
meanQ19	.6730907	1.07198	0.63	0.531	-1.438339	2.78452
meanQ20	.4591864	.8385996	0.55	0.584	-1.192565	2.110938
meanAgency1City1	.100029	.6890203	0.15	0.885	-1.257103	1.457161
meanAgency2City1	.5779752	.6178098	0.94	0.350	-.6388964	1.794847
meanAgency1City2	.5055883	.9991751	0.51	0.613	-1.462441	2.473618
meanAgency2City2	2.190931	1.075331	2.04	0.043	.0729008	4.30896
meanAgency1City3	1.378618	.7396083	1.86	0.064	-.0781547	2.835391
meanPFSAge	.2259529	.1976697	1.14	0.254	-.1633881	.615294
meanAge1st_arr	-.0452084	.1864719	-0.24	0.809	-.4124936	.3220767
meanArr_pre_total	-.2027236	.102997	-1.97	0.050	-.405592	.0001449
meanInc_pre_total	-.0249239	.2336046	-0.11	0.915	-.4850441	.4351963
meanPreDays	-.0001983	.0007921	-0.25	0.803	-.0017584	.0013619
meanPreEmpl	1.291363	.1816417	7.11	0.000	.9335915	1.649134
TxAssigned	-.1158614	.2781176	-0.42	0.677	-.6636569	.4319341
_cons	-.5013105	4.359824	-0.11	0.909	-9.088657	8.086036

Notes: The key term in the model is TxAssigned. (The covariates for the employment outcome regressions are means for those values across groups of 6 to 11 individuals.)

Exhibit 26. RCT IV estimate for Employment (meanPostEmpl)

```
Instrumental variables 2SLS regression          Number of obs    =        278
                                                Wald chi2(31)    =     190.16
                                                Prob > chi2      =     0.0000
                                                R-squared        =     0.4048
                                                Root MSE         =      1.905
```

meanPostEmpQ40	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
TxEnrolled	-.3974306	.9010492	-0.44	0.659	-2.163455	1.368593
meanQ2	1.809465	.8069896	2.24	0.025	.2277949	3.391136
meanQ3	2.245306	1.10251	2.04	0.042	.0844252	4.406186
meanQ4	1.758559	.9038533	1.95	0.052	-.0129606	3.530079
meanQ5	1.079283	.7332965	1.47	0.141	-.3579521	2.516517
meanQ6	2.023633	.7346057	2.75	0.006	.5838327	3.463434
meanQ7	1.346302	.8292504	1.62	0.104	-.2789991	2.971603
meanQ8	1.373389	1.055048	1.30	0.193	-.6944666	3.441245
meanQ9	1.963539	.9023863	2.18	0.030	.1948948	3.732184
meanQ10	1.455283	.8352396	1.74	0.081	-.1817566	3.092322
meanQ11	1.370619	1.029023	1.33	0.183	-.6462297	3.387467
meanQ12	1.966054	.7485161	2.63	0.009	.4989892	3.433119
meanQ13	1.67992	.9328508	1.80	0.072	-.148434	3.508274
meanQ14	1.641818	.8222065	2.00	0.046	.0303228	3.253313
meanQ15	-.5011183	.9395718	-0.53	0.594	-2.342645	1.340409
meanQ16	1.207644	.8585337	1.41	0.160	-.4750507	2.89034
meanQ17	.8841129	.9312512	0.95	0.342	-.941106	2.709332
meanQ18	2.227992	.8909186	2.50	0.012	.481824	3.974161
meanQ19	.7589336	1.056652	0.72	0.473	-1.312066	2.829933
meanQ20	.5123446	.8319866	0.62	0.538	-1.118319	2.143008
meanAgency1City1	.1083222	.6490781	0.17	0.867	-1.163847	1.380492
meanAgency2City1	.5461859	.5867458	0.93	0.352	-.6038148	1.696187
meanAgency1City2	.5125769	.9416188	0.54	0.586	-1.332962	2.358116
meanAgency2City2	2.212276	1.022277	2.16	0.030	.2086498	4.215902
meanAgency1City3	1.402487	.7080441	1.98	0.048	.0147463	2.790228
meanPFSAge	.1895908	.1854137	1.02	0.307	-.1738134	.5529951
meanAge1st_arr	-.0440364	.1759221	-0.25	0.802	-.3888373	.3007646
meanArr_pre_total	-.2089642	.0996925	-2.10	0.036	-.404358	-.0135704
meanInc_pre_total	-.0448206	.2361039	-0.19	0.849	-.5075757	.4179346
meanPreDays	-.0000883	.0008185	-0.11	0.914	-.0016926	.001516
meanPreEmpl	1.265247	.1798584	7.03	0.000	.9127307	1.617763
_cons	.3086479	4.211544	0.07	0.942	-7.945827	8.563123

Instrumented: TxEnrolled

Notes: The key term in the model is TxEnrolled. The covariates for the employment outcome regressions are group means for those values.

APPENDIX C: DID PROPENSITY SCORE MODELS

Contemporaneous Comparison Group

```
Logistic regression                                     Number of obs =  1,418
                                                        LR chi2(10)   =  68.67
                                                        Prob > chi2   = 0.0000
Log likelihood = -923.83048                             Pseudo R2     = 0.0358

--------------------------------------------------------------------------------------
         TvC_A_juris | Odds ratio   Std. err.      z    P>|z|     [95% conf. interval]
---------------------+----------------------------------------------------------------
persTotDaysFDISP_pre |   1.000228   .0001155     1.97   0.049     1.000001    1.000454
 persNumCasesARR_pre |    .987304   .0161822    -0.78   0.436     .9560914    1.019535
 persNumCasesINC_pre |   .8713298   .0342205    -3.51   0.000     .8067752    .9410496
     preNum_low_pers |   .9912438   .1402373    -0.06   0.950     .7512005    1.307992
     preNum_med_pers |   .9030153   .0849403    -1.08   0.278     .7509803     1.08583
   preNum_medhi_pers |   1.610528   .2520742     3.04   0.002     1.185062    2.188746
    preNum_high_pers |   .6770758   .3263569    -0.81   0.418     .2632389    1.741504
preNum_violDrug_pers |   1.656214   .2152758     3.88   0.000     1.283739    2.136762
     preEmpl_perPers |   .5854164   .2445105    -1.28   0.200     .2581928    1.327351
        ageAt_1stArr |   1.038792   .0259814     1.52   0.128     .9890973    1.090983
               _cons |   .5169447   .2495519    -1.37   0.172     .2006928    1.331547
--------------------------------------------------------------------------------------
Note: _cons estimates baseline odds.
```

Note: The predictors are:

```
persTotDaysFDISP_pre: total days of previous  incarceratio
 persNumCasesARR_pre: number of previous arraignments
 persNumCasesINC_pre: number of previous incarcerations
     preNum_low_pers: number of previous arraignments on low-level offenses
     preNum_med_pers: number of previous arraignments on medium-level offenses
   preNum_medhi_pers: number of previous arraignments on medium-high-level offenses
    preNum_high_pers: number of previous arraignments on high-level offenses
preNum_violDrug_pers: number of previous arraignments on violent or drug offenses
     preEmpl_perPers: number of prior months of incarceration
        ageAt_1stArr: age at first arrest
```

Historical Comparison Group

Logistic regression | Number of obs = 1,693
LR chi2(12) = 239.94
Prob > chi2 = 0.0000
Log likelihood = -966.21495 | Pseudo R2 = 0.1105

TvC_B_histRoca	Odds ratio	Std. err.	z	P>|z|	[95% conf. interval]	
persTotDaysFDISP_pre	1.00059	.0001383	4.26	0.000	1.000319	1.000861
persNumCasesARR_pre	1.06113	.0192457	3.27	0.001	1.024072	1.099529
persNumCasesINC_pre	.9938783	.0417782	-0.15	0.884	.9152769	1.07923
preNum_low_pers	.7080603	.0927035	-2.64	0.008	.5478049	.9151971
preNum_med_pers	.9565075	.0834171	-0.51	0.610	.8062226	1.134806
preNum_medhi_pers	1.001458	.1442065	0.01	0.992	.7552016	1.328015
preNum_high_pers	.356247	.1865353	-1.97	0.049	.127659	.9941473
preNum_violDrug_pers	1.329774	.1683663	2.25	0.024	1.03754	1.704319
preEmpl_perPers	98.60321	43.24927	10.47	0.000	41.73858	232.9402
ageAt_1stArr	1.216382	.0357719	6.66	0.000	1.148253	1.288554
city_DID						
2	1.057541	.1526735	0.39	0.698	.7969153	1.403402
3	1.510041	.2052276	3.03	0.002	1.15692	1.970944
_cons	.005905	.0033214	-9.12	0.000	.0019608	.0177829

Note: _cons estimates baseline odds.

Diagonal Comparison Group

Logistic regression | Number of obs = 1,264
LR chi2(10) = 192.67
Prob > chi2 = 0.0000
Log likelihood = -775.00841 | Pseudo R2 = 0.1106

TvC_C_diag	Odds ratio	Std. err.	z	P>|z|	[95% conf. interval]	
persTotDaysFDISP_pre	1.001012	.0001764	5.74	0.000	1.000667	1.001358
persNumCasesARR_pre	1.087662	.021717	4.21	0.000	1.045919	1.13107
persNumCasesINC_pre	.8976519	.041993	-2.31	0.021	.8190076	.983848
preNum_low_pers	.7939179	.1279445	-1.43	0.152	.5788947	1.088809
preNum_med_pers	.9957897	.106767	-0.04	0.969	.807055	1.228661
preNum_medhi_pers	1.337838	.2334629	1.67	0.095	.9503037	1.883409
preNum_high_pers	.6221266	.348078	-0.85	0.396	.2077945	1.862617
preNum_violDrug_pers	1.544218	.2208669	3.04	0.002	1.166709	2.043876
preEmpl_perPers	26.75411	12.55352	7.00	0.000	10.66575	67.11042
ageAt_1stArr	1.235259	.0362872	7.19	0.000	1.166146	1.308468
_cons	.0082034	.0046177	-8.53	0.000	.0027218	.0247248

Note: _cons estimates baseline odds.

APPENDIX D: DID REGRESSION MODELS

Exhibit 27: DID ITT estimate on incarceration days

```
(sum of wgt is 2,652.89982882328)

Linear regression                               Number of obs     =      3,064
                                                F(3, 3060)        =       1.49
                                                Prob > F          =     0.2151
                                                R-squared         =     0.0024
                                                Root MSE          =     641.56

--------------------------------------------------------------------------------
                |               Robust
      RecidDays | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
----------------+---------------------------------------------------------------
 DESIGNperiodTx |  -54.71866   52.45991    -1.04   0.297    -157.5789    48.14156
DESIGNjurisRoca |   7.465141   43.14117     0.17   0.863    -77.12345    92.05373
       DIDtxEff |  -16.93466   65.49543    -0.26   0.796    -145.3541    111.4848
          _cons |   365.3858    33.0272    11.06   0.000     300.6281    430.1435
```

Note: The key term is DIDtxEff, which is the interaction term between the periods (contemporary vs. historical) and the jurisdictions (Roca-served areas vs. comparison).

Exhibit 28: DID ITT estimate on employment

```
((sum of wgt is 2,652.89982652292)

Linear regression                               Number of obs     =        824
                                                F(3, 820)         =       0.54
                                                Prob > F          =     0.6524
                                                R-squared         =     0.0026
                                                Root MSE          =     2.6217

--------------------------------------------------------------------------------
                |               Robust
postEmpl_perP~A | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
----------------+---------------------------------------------------------------
 DESIGNperiodTx |   .0597613   .4403446     0.14   0.892    -.8045739    .9240966
DESIGNjurisRoca |  -.2260854   .3815903    -0.59   0.554    -.9750943    .5229234
       DIDtxEff |    .247796   .5140974     0.48   0.630    -.7613058    1.256898
          _cons |   5.360014   .3401744    15.76   0.000     4.692299     6.02773
--------------------------------------------------------------------------------
```

Note: 1) The key term is DIDtxEff, which is the interaction term between the periods (contemporary vs. historical) and the jurisdictions (Roca-served areas vs. comparison.)

2) The unit of analysis was the employment group. Employment data was requested for groups of 6 to 11 individuals, and results were returned as group totals (and then divided by the group size, so as to be scaled for individuals). Because group sizes varied, group results were weighted by group size.

APPENDX E: RESULTS ACROSS PROJECT ESTIMATES

Incarceration Days

The effects of random assignment (ITT estimate) and enrollment (IV estimate) were estimated 13 times during the course of the project. Especially in early estimates, the regression estimates were particularly important because uncontrolled comparisons tend to be systematically biased toward larger outcomes for the treatment group, as the result of more individuals having been intentionally randomized to treatment in early project quarters (see Appendix A). This gave those early individuals longer observed periods of follow-up, in turn allowing them greater opportunity to accrue outcomes, whether detrimental (i.e., being sentenced to incarceration) or beneficial (i.e., employment).

Exhibit 29 displays how the incarceration estimates changed over the course of the project. Each new estimate added new randomization cohorts, and also lengthened the follow-up observation period for all observed participants (until it reached 5 years). ITT estimates are displayed in the top panel, and the IV estimates in the bottom panel.

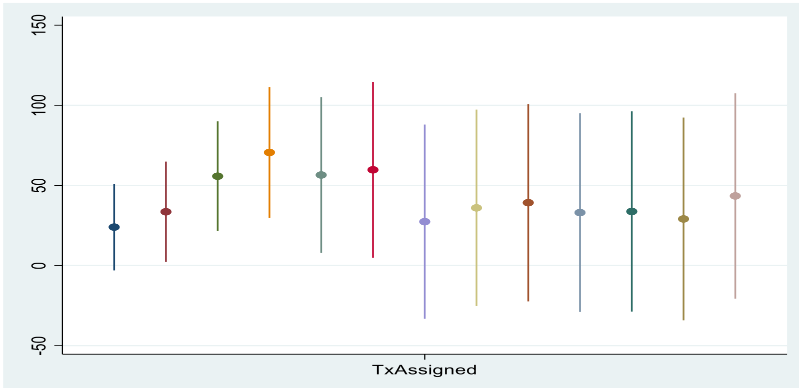
Several things are notable: First, the point estimates have consistently been in the direction of more incarceration days for individuals randomized to treatment. Second, results have been close to null except for the 3rd through 6th estimates, which were larger, but then came back down. Third, despite growing sample sizes, the confidence intervals across the estimates also increased for the first six estimates, which is somewhat surprising. By the 7th estimate, results had stabilized.

It is important to realize that these estimates are *not* independent; most of the people observed and most of the time observed does not change that much from estimate to estimate (although the proportion of new people observed is larger with earlier than later estimates). Stability of non-independent estimates should not be confused with their reliability in a statistical sense (or statistical significance), which is better illustrated by the confidence-interval bars.

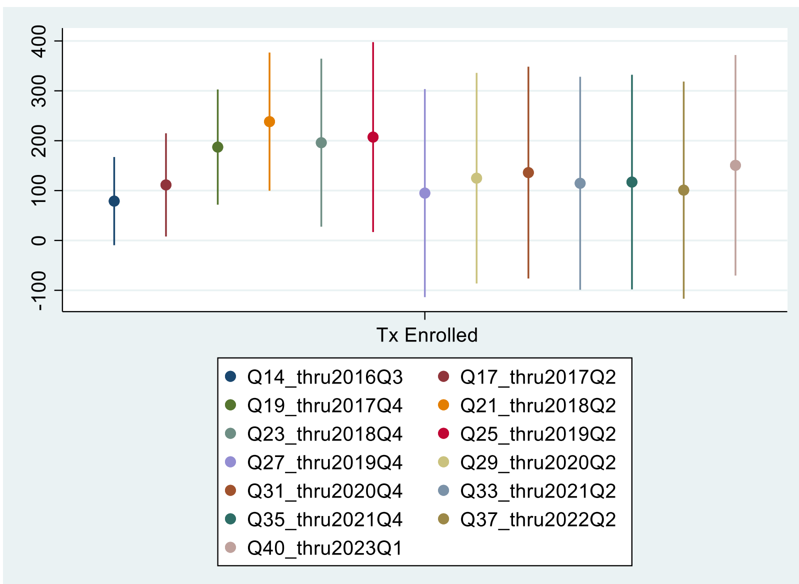
Exhibit 30 displays the odds ratio over the course of the project. The point estimates have consistently been in the direction of higher odds of incarceration for individuals randomized to treatment (odds ratios greater than 1.0), but these results have been close to null except for the 1st, 3rd and 4th estimates, and estimations had stabilized by the 5th estimate. And in contrast to the incarceration days results, here the confidence intervals get smaller with the growing sample sizes of later estimates.

Exhibit 29. RCT Incarceration Days Results Across Project Estimates

ITT ESTIMATES

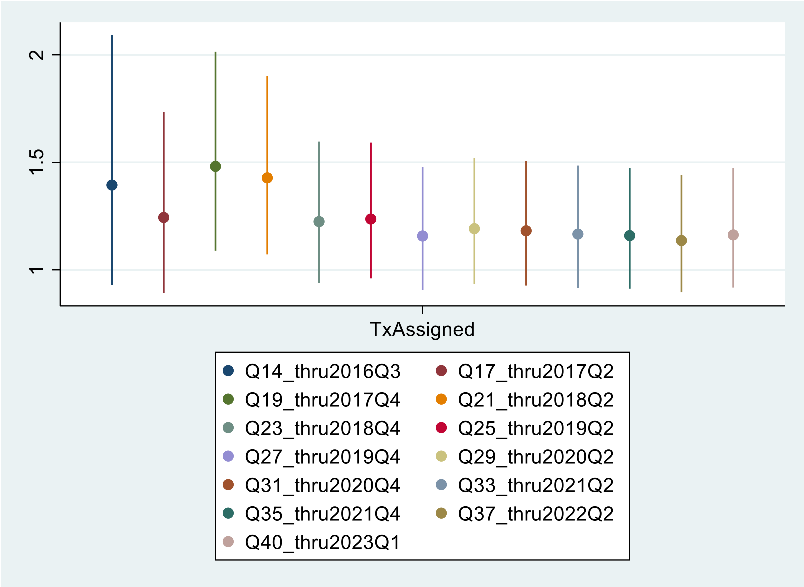


IV ESTIMATES



Note: Dots show point estimates and bars show 95 percent confidence intervals.

Exhibit 30. RCT Logistic Regression Odds Ratio Results Across Project Estimates



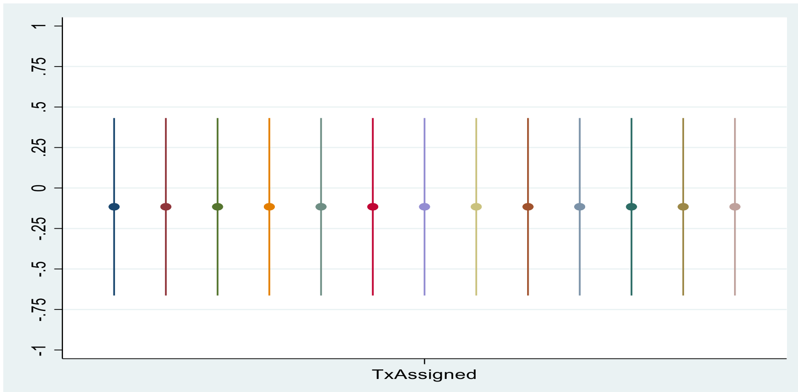
Notes: 1.0 is a null result. Dots show point estimates and bars show 95 percent confidence intervals. Because these estimates are not independent, their stability should not be confused with their statistical reliability or significance, which is better illustrated by the confidence-interval bars.

Employment Estimates

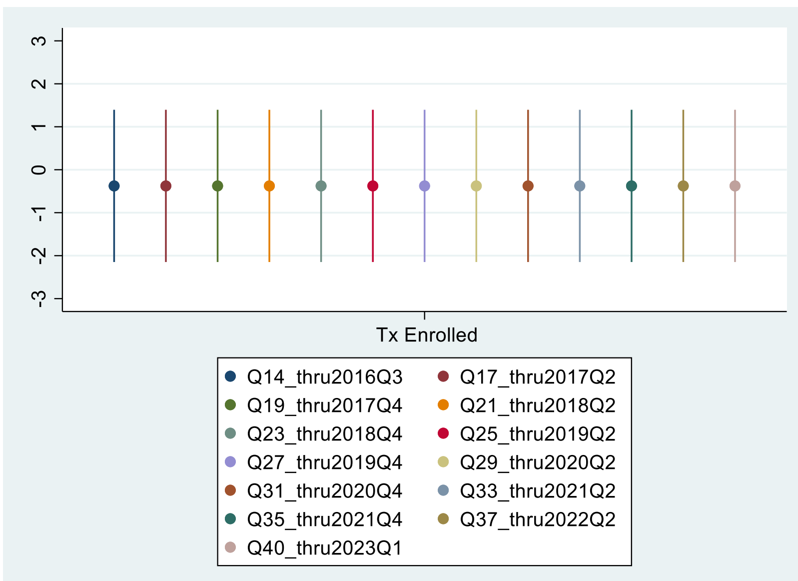
Finally, Exhibit 31 displays how the employment results have changed over time with subsequent estimates, showing the ITT estimates in the top panel, and the IV estimate results in the bottom panel. Each new estimate added new randomization cohorts to those in the prior estimate; the relevant randomization cohorts are shown in the legend. Each new estimate also lengthened the follow-up observation period for all observed participants (until it reaches 5 years). The point estimates for employment were in the direction of less employment for individuals randomized to treatment, but close to null and with 0 within the confidence intervals.

Exhibit 31. RCT Results for Employment Across Project Estimates

ITT estimate



IV estimate



Note: Dots show point estimates and bars show 95 percent confidence intervals. Because these estimates are not independent their stability should not be confused with their statistical reliability or statistical, which is better illustrated by the confidence-interval bars.

APPENDIX F: THE REVISED EVALUATION PLAN

REVISED EVALUATION PLAN
FOR EXTENDED ROCA EVALUATION

August 4, 2020, rev. April 30 2023 (minor corrections Aug. 8, 2023)

Akiva Liberman
Independent Evaluator

This evaluation plan for the extended MA PFS contract describes both the two RCT estimates (ITT and IVE) and the Difference in Difference (DID) estimate. These will then then be combined to generate a "backstop estimate." Except as noted, this plan does not change the prior plan, but rather is primarily intended to memorialize the plan and any decisions that were made for implementing the backstop methodology and extending the evaluation.

- Because randomization is complete, which ran from 2014 through the end of 2018, this document does not discuss the details of the randomization process.

- Details from prior documents are included in Appendices, as shown in the Table of Contents on the next page.

Since 2020, this document has been revised based primarily on four sources:

(a) My responses to comments/questions on the 2020 draft;

(b) Updated sample sizes with those from my 2/20/2023 memo on sample sizes;

(c) Discussion of design modifications from my DID Design Considerations memo, revised, dated 3/7/2022;

(d) My improved understanding of the risk scores, and particularly the HKS risk score.

- In the RCT, the HKS risk score was used only for non-DOP referrals -- which are excluded from the DID estimate.

- Aside from consideration of jurisdiction (Boston vs. Chelsea vs. Springfield), the HKS risk score was based only on criminal history variables (shown in Exhibit 4 in the discussion of propensity score variables).

Table of Contents

6/1/24, page 2

I. OVERVIEW

The Analytic Samples

A. Between 2014 and 2018, 1,812 eligible young men aged 17-24 who were released from incarceration in the Boston, Chelsea, or Springfield areas were randomized to Roca or control, in a randomized controlled trial (RCT).

B. These young men were referred by the Department of Probation (DOP; n=1,147) or only by other state or local agencies (n=665).

C. Among those who were referred, Roca successfully made initial contact with 446 of these men, who were then considered "enrolled."

D. Roca also simultaneously recruited other individuals for the program. Among individuals who had been assigned to the control group, 54 enrolled in the program (sometimes referred to as Roca "self-recruits").

All of the preceding samples are involved in the RCT estimates. In addition, a difference in difference (DID) estimate is being conducted. The treatment sample for the DID estimate involves just referrals to randomization from DOP. It also involves three additional comparison samples:

E. Similar young men in the program period (2014-2018), released to DOP but in comparison jurisdictions in MA (Worchester, Lawrence, Fall River, Brockton).

F. Similar young men released to DOP during a previous period (2010-2012), in the jurisdictions used in the RCT.

G. Similar young men released to DOP during a previous period (2010-2012), in the comparison jurisdictions.

Exhibit 1 illustrates how these samples are related to each other. Cells with a red font are involved in the RCT estimates; blue cells are involved in the DID estimates.

Final Samples. The RCT samples of those who are considered enrolled – or not enrolled -- in Roca are now finalized to the samples shown above. (Randomization ended in 2018. Through 2019, the IE requested new enrollment lists from Roca for each estimate, and the enrollment samples were therefore somewhat different in each estimate.) Therefore, any changes in outcomes during the extended evaluation period will be due solely to changes in the outcomes, rather than any changes to the samples.

- In early 2020, the Urban Institute (UI) was able to identify all of the eligible individuals in the comparison jurisdictions during the program period, and exclude people based on the same criteria used for the RCT, which are attached in APPENDIX 2. After exclusion, a sample was identified of 607 comparison individuals.

6/1/24, page 3

ATTACHMENT A

Exhibit 1. Samples for Evaluation

		ENROLLMENT						RCT JURISDICTIONS			COMPARISON JURISDICTIONS
		Not Enrolled			Enrolled			Random Assignment			
		C	T	Total	C	T	Total	C	T	Total	
RCT Sample	All Referral Sources	572	740	1,312	54	446	500	626	1,186	1,812	
	DOP Referrals	**343**	**491**	**834**	**29**	**284**	**313**	**372**	**775**	**1,147**	
	Non-DOP Referrals	229	249	478	25	162	187	254	411	665	
Other DID-specific exclusions	admin probation	79	95	174	4	20	2	83	115	198	
	2012 probation start	15	36	51	0	16	16	15	52	67	
	age 17 referral	1	0	1	1	0	1	2	0	2	
	Total	**95**	**131**	**226**	**5**	**36**	**19**	**100**	**167**	**267**	
DID Samples	**RCT cases eligible for DID (DOP referrals minus other DID-specific exclusions)**		360			248			608		607
	Historical Period (2010-2012), DOP Referrals									1,457	926

Note: C refers to control group, T to treatment group

6/1/24, page 4

Estimates of Program Impact

- The Intent to Treat Estimate (ITT), based on the RCT, examines the effect of random assignment, based on the individuals in A above -- regardless of enrollment.

- The Instrumental Variable Estimate (IVE), also based on the RCT, estimates the effect of enrollment in Roca, by taking into account the individuals in C and D above, who were randomly assigned to Roca but did not enroll[1] or who enrolled without being randomly assigned. (These individuals are sometimes called "non-compliant" with their assigned condition.)

- In the DID estimates, the treatment group consists only of individuals randomized to Roca after referral from DOP.

- For the DID estimates, we also want to produce estimates of the effect of enrollment, controlling for people who were assigned to treatment but did not enroll (group C), We refer to this again as an IVE, but the method is slightly different than for RCT estimates, because there are no people who enrolled among the comparison groups. For the DID, the IVE is simply a multiplication of the ITT estimate by the ratio of those assigned to treatment who actually enrolled, which now can be seen as being 2.73 = 775 / 284.[2]

- The DID and RCT estimates will then be combined to create the "backstop estimate," with weighting as recommended by Jeffrey Leibman at HKS. The basic criterion was to weight the estimates in inverse proportion to their expected standard errors. In the original plan, the RCT was to be weighted at 80% and the DID at 20%. In 2019 Jeffrey Leibman at HKS recommended changing the weights to 70/30, based on new simulations with the final RCT sample sizes. Now, with the final DID samples in hand, Professor Leibman will be providing updated weights.

- The backstop IVE estimate will be the basis for recidivism and employment success payments.

Outcomes

Recidivism Outcomes

Recidivism outcomes are the primary outcome for the continued evaluation. The recidivism outcome is the number of sentenced days of incarceration (aka "bed days") on new charges – excluding technical violations, within 5 years since the beginning of follow-up. Bed days are capped at 10 years of sentences. The follow-up period begins the day after randomization for the RCT samples. Analyses of the RCT data concerning time from the beginning of probation

---

[1] Inability to reach individuals with incorrect address information led some of the young men who were randomized to Roca to be designated as "nonconforming referrals." These individuals are retained in the ITT. Because Roca did not try to enroll individuals in the control group, equivalent individuals with bad addresses could not be excluded from the control group. However, because these people never enrolled, the IVE corrects for this issue.

[2] This was described in Backstop Methodology, p. 6.

until randomization suggested that the observation period should begin one year after entry into probation for the DID comparison groups (see DID Design Considerations Memo, I.3).

Employment Outcomes

The employment outcome is defined as the number of eligible quarters in which an individual has earnings greater than or equal to $1,000 in the Massachusetts (MA) unemployment insurance (UI) database.

Employment outcomes are the secondary outcome for the continued evaluation. For confidentiality reasons, DUA does not return employment data on identified individuals, but agreed to return data aggregated to groups consisting of at least 6 individuals. Therefore, the sample is organized into groups of at least six people each, and those groups are then sent to DUA for employment outcome data, which is returned at the group level.

For the DID estimate, which is limited to Probation referrals, we must ensure that Probation individuals in the treatment sample are not grouped with others. Therefore, the grouping procedure for all samples – including both RCT treatment and comparison samples – will create groups of individuals that are each composed only of either Probation or non-Probation referrals. (This was a slight change to the prior grouping procedure, which was adopted with the Q25 estimate). The particulars of grouping the sample are described in APPENDIX 1.

Imputation of Bed Days for the RCT samples at the final estimate only.

The goal of the extended evaluation is to approach the intended 5-year follow-up period for all individuals. For the historical comparison groups, 5 years have already elapsed. For the RCT samples, randomization was conducted through Dec. 2018.

1) At the time of the final estimate, with data through March 2023 (Q38), all individuals except those randomized in the last 3 quarters will have complete outcome data. For those individuals randomized in the last 3 quarters of 2018, bed day outcomes through 60 months will be imputed using the imputation process developed by HKS in the original evaluation contract.

2) The prior evaluation plans also call for imputation of bed days for individuals with open arraignments at the time of the final estimate, as a function of the top charge in that open arraignment.

The imputation processes from the original evaluation contract are included in APPENDIX 3.

Modifications Considered

We note here several changes that were discussed concerning the estimation methodology, but which are not being adopted for payment purposes, although they might be explored for sensitivity analyses. These include (a) dropping Q2 cases, (b) reducing the maximum sentence from 10 years to 5 years, and (c) starting the follow-up period 90 days after randomization rather than on the date of randomization.

6/1/24, page 6

Schedule of Estimates

The schedule of estimates is shown in the table below.

Exhibit 2. Schedule of Evaluation Estimates (from SOW in Akiva Liberman LLC evaluation contract).

Child Trends Estimate	Outcomes observed through	End of Quarter	Quarter in which outcome data is requested aka Outcome Measurement Quarter (OMQ) (allowing 6 months lag for data accrual)	End of Quarter	Draft Estimate Due	End of Quarter
1	PFS Q27	June 2020	PFS Q29	Dec 2020	PFS Q31	June 2021
2	PFS Q29	Dec 2020	PFS Q31	June 2021	PFS Q33	Dec 2021
3	PFS Q31	June 2021	PFS Q33	Dec 2021	PFS Q35	June 2022
4	PFS Q33	Dec 2021	PFS Q35	June 2022	PFS Q37	Dec 2022
5	PFS Q35	June 2022	PFS Q37	Dec 2022	PFS Q39	June 2023
Final	PFS Q38	Mar 2023	PFS Q40	Sept 2023	PFS Q42	Mar 2024

Payment Determination

Success payments are based upon the IVE estimate of the effect of enrollment, using the backstop methodology to combine the RCT and DID estimates. If this estimate shows a benefit from program enrollment, in terms of reduced bed-days and/or and increase in the number of months of employment over $1,000, then the IE will estimate payments due to Roca, based on the schedule in the Second Amended PFS contract. The payment provisions are summarized here:

- The Independent Evaluator shall calculate the "PFS Payments Earned to Date Due to Gains in Employment", which shall be equal to the Total Employment Gain multiplied by $750.
- The payment formula for bed days avoided is shown in the table below. For the purposes of this calculation the Bed Days Avoided Per Person Served shall be rounded to the nearest hundredth of a bed day.

6/1/24, page 7

Bed Days Avoided Per Person Served[*]	**Payment Formula**
>=0 and <29	no payment
>=29 and <88	$785 + ((Bed Days Avoided Per Person Served – 29) x 55)
>=88 and <244	$4,016 + ((Bed Days Avoided Per Person Served – 88) x 145)
>=244 and <359	$26,639 + ((Bed Days Avoided Per Person Served – 244) x 16)
>=359	$28,540

- Because all estimates of benefits are cumulative, prior payments are subtracted from the total due at any time.

- Except for the final payment, along with each estimate, 20% of the payment estimates are held back so that only 80% of the estimated payments are due.

- Average payments *per person* will be estimated due to estimated reductions in recidivism or increases in employment. These per person payments will then also be due for all ”Roca Recruits”, also known as “self-recruits”. (Because outcomes are based on administrative data only, individuals randomly assigned to the control group were unaware of the study, nor did Roca know their identities.)

II. OBTAINING DATA

Exclusions from DID samples

For the RCT samples, once the appropriate age and period samples are identified, data requests were made to DCJIS and Roca to obtain data used for exclusion (see APPENDIX 2. EXCLUSION CRITERIA). With the samples finalized, there is no longer a need to obtain additional Roca data for exclusion.

Contemporaneous Comparison Group Data

At present, DOP has already sent data to Urban for the DID comparison groups. For the contemporaneous comparison group, the data sent to Urban during randomization was revisited. (That referral data had not been limited to the Roca jurisdictions; cases from eligible jurisdictions were then identified for randomization.) Urban processed this data in early 2020, selecting just the cases from the comparison jurisdictions and running the exclusion criteria that had been used for RCT.

- *However, the DID involves using the individual variables used in the HKS risk score (see discussion of the Backstop Estimate) for creating propensity scores (PSs) for the historical comparison groups, because those groups did not have the same calculated risk scores as the cases in the contemporaneous period.*

Historical Comparison Group Data

For the historical comparison groups, data was received in early 2020 and has not yet been processed, but we anticipate that it will be broadly sufficient.

- *However, the DOP data sets for the historical sample did not include SSN, which is the key identifier for the employment outcome data. Once the historical sample is identified, the list of individuals (names, DOB, pcf number) will be sent to DCJIS – along with the request for CJ outcome data -- and SSNs requested, before employment data can be obtained request from DUA.*

Recidivism Outcome Data

Information used to calculate bed days of recidivism will be obtained by the IE from the MA Department of Criminal Justice Information Services (DCJIS) Criminal Record Offender Information (CORI) database. The IE will send to DCJIS the list of PCF numbers and other identifying information (Names, DOBs, SSNs) for the individuals on the measurement data file for automated matching to CORI data via the Commonwealth's secure Interchange system.

Employment Outcome Data

The MA Department of Unemployment Assistance (DUA) will provide quarterly UI earnings data to the IE aggregated into groups of six or more individuals. Accordingly, the IE will assign individuals to groups prior to requesting DUA data.

The IE will send to DUA via secure electronic file transfer all individuals in the measurement data file that have a 9-digit Social Security Number (SSN) obtained prior to randomization. This file will contain the SSNs and corresponding Group Identification (Group ID) numbers.

- The RCT data does not consistently contain SSN data, which will be obtained from DCJIS.

III. RCT ESTIMATES

Covariates

In the RCT analyses, the same covariates are used for both recidivism and employment outcomes. Because employment outcomes obtained for small groups of individuals (6 to 11 people), the analyses are of those groups, and the covariates are the means for those groups.

- Quarter of randomization
- Agency source type interacted with geographic service area
- Age at randomization
- Age at first arraignment
- Number of pre-randomization arraignments
- Number of pre-randomization incarcerations
- Days of pre-randomization incarcerations
- Pre-randomization employment calculated as the number of the prior eight quarters in which an individual's UI earnings were greater than or equal to $1,000. (These data are only available for small groups, rather than individuals; see employment outcome, above.)

The regression specifications for both the ITT and IVE for both outcomes are included in APPENDIX 4.

6/1/24, page 10

IV. BACKSTOP ESTIMATE

In light of lower enrollment numbers than hoped, the PFS contract and evaluation contract called for statistical analyses to create backstop estimates of program impact, through combining the RCT estimates with DID estimates.

In Jan 2020, Jeffrey Leibman at the Harvard Kennedy School (HKS) conducted additional simulations based on the sample sizes in the actual RCT (rather than projected sample sizes which were used in the original plan). Professor Liebman will make a final determination of the appropriate weighting prior to the calculation of the DID.

Difference in Difference (DID) Analysis and Comparison Groups

A DID analysis is two dimensional. Here one dimension is historical time;[3] the other dimension is differences in jurisdictions. It is the intersection of these two dimensions (a two-way interaction in regression analyses) that is key to assessing effects.

The DID requires data for four samples, as illustrated in Exhibit 3 (also see Exhibit 1 earlier), The treatment sample is from the RCT, along with three new comparison groups.

Exhibit 3: Difference in Difference Design

	A. Historical differences for treatment group	**B. Young men released to other jurisdictions: "Same age and risk, different geography"**	**Differences Across Jurisdictions**
1. Program period (released or on probation 2013-2018)	A1. Individuals who were randomized to Roca (data from RCT)	B1. Excepting jurisdictions, same requirements as A1	A1-B1: Cross-jurisdiction difference
2. Historical period (2010-2012)	A2. Prior cases meeting the eligibility requirements for the RCT	B2. Excepting jurisdictions, same requirements as A2	A2-B2: Historical Cross-Sectional Difference
Differences over Time	A1 – A2: "historical difference"	B1 – B2: Difference over time for similar young men in other jurisdictions	**Difference in Differences** = (A1-B1)-(A2-B2) = (A1-A2)-(B1-B2)

Time Period for Comparison Samples

Consistent with the time period of the randomization, the contemporaneous comparison group involves data on young men on released to parole from 2013 through 2018. (Although

---

[3] "Historical time" is intended to indicate that different cohorts, from different periods, are being compared – rather than behavioral change over time for the same group of people. Both are sometimes referred to as "pre-post" comparisons.

randomization began in 2014, it drew on individuals already on parole or probation, many of whom had been released in 2013.)

To prevent the samples from overlapping, the historical comparison sample consists of individuals on released to parole in 2010-2012 in the treatment and comparison geographic areas.

Overlapping Samples due to Recidivism

Some individuals in the historical groups may have recidivated and then have been identified in the program period. These people will *not* be excluded, because excluding them could systematically bias some samples against recidivists. For example, taking the treatment group (A1) as given, we might consider excluding those same individuals from A2. But this would seem to systematically exclude a set of individuals who fail during period A2, which is how they end up released again during the period for A1.

- The historical comparison group consists of individuals who were released from incarceration to DOP from 2010-2012. Some of these will then have participated in Roca in 2011-2012, during a period which is treated as an exclusion criterion for the RCT (which began in 2014). We will want to identify those Roca participants, through a data request from Roca, and treat those individuals as “enrolled” for the DID’s instrumental variable estimate. (Note that, as discussed above, we will *not* exclude from the historical comparison groups those individuals who then end up in the RCT, because that would bias the comparison samples.)

Referral Sources

The RCT involves referrals from both state and local (Sheriffs) referral sources. The backstop will only involve individuals on Probation (both "Admin Probation" and "MA Office of Probation"), which comprise the majority of the referrals to the RCT. For the DID, the treatment group will also be limited to that same referral source, as shown earlier in Exhibit 1.

Target Sample Size for New Datasets

The evaluation plan says “… to reach a sample size roughly equivalent to that of the RCT” (fn. 2, p. 4). We interpreted this to mean a rough target sample size for each comparison groups of approximately 765, which was the hoped-for size of the enrolled treatment group, rather than the actual size of the treatment group as either referred (over 1000) or enrolled (about 400).

DID sample sizes, shown above in Exhibit 1, are as follows:

	Comparison cities	Roca cities
CONTEMPORARY GROUPS	607	608
Historical groups	926	1,457

Analyses

Outcome Observation Period and Imputation – for DID estimated at Q38

The goal of the evaluation is to estimate the effects of Roca on outcomes over 5 years. For the RCT samples, the follow-up time begins on the day after randomization. As discussed on p. 3, based on analyses of RCT data concerning time from the start of probation until randomization,

6/1/24, page 12

the observation period for the DID comparison groups will begin 1 year following the start of probation.

For the historical samples (A2 and B2), we will obtain follow-up observation data for five years. In the DID design, a small number of individuals in the program period will have shorter observation periods. At the recommendation of Jeffrey Leibman no imputation will be conducted for recidivism outcomes,[4] because these differences in observation period will be implicitly controlled by the DID design.

Propensity Score Weighting

The backstop plan calls for weighting by propensity scores (PSs), to insure comparable samples. Propensity scores come from equations that summarize the difference between two groups. They are then used in some quasi-experimental designs to remove differences between the groups (a.k.a., controlling the "case mix"). In this case, a separate PS will be computed for each of the three comparison samples (A2, B1, B2) compared to the randomized treatment group (A1). Weights for the comparison samples will then be computed from that comparison group's PS, so that each comparison group is as similar to the treatment group as possible, As is typical in estimating the average treatment on the treated (ATT), treatment individuals receive weights of 1.0.

Trimming

Cases for the comparison groups will be excluded ("trimmed") if they either (a) have propensity scores below any found in the program period, or (b) if the PS weight for any cases is below 0.20 or above 5.0.

Propensity Score Variables

Quarter of randomization – this will be included just in the PSs for the contemporaneous comparison group. Because PS weighting leads to samples that are balanced on all included covariates, this will control for historical changes during the period of randomization and generate a weighted comparison sample distributed over time like the treatment group.[5] This will accomplish what is done through weighting in the RCT estimate.

- We will include the variables that comprise the HKS risk scores per se in creation of the PSs. However, the HKS risk score included dummy variables for the 3 jurisdictions in the RCT. Although these cannot be used for PSs for the comparison groups in non-RCT jurisdictions, we will include these variables when creating the PS for the historical group in the RCT jurisdictions (i.e., group F on page 1).

- Other covariates that are used in the RCT will also be controlled, either in the PS or simply as covariates.

---

[4] No imputation is contemplated in the contract for observation windows shorter than 5 years, perhaps because historical analyses found the employment outcome to be relatively flat over time, as seen in Figure 1 of the technical narrative for the DOL grant.

[5] With 5 years of randomization, this involves 5 x 4 -1 = 19 dummy variables. By chance, we would expect an average of 1 of those to be "unbalanced." Balance over the entire set can be checked through a 2-way chi-square (Quarter by sample).

- Although education data was mentioned in the Backstop Methodology document, they will not be included. To be usable for current purposes, those data would need to include dates of graduation and/or GEDs, which is not available.
- Although the original backstop methodology document (p. 5) mentioned arrest, we are using arraignment (which I expect was the original intention).

6/1/24, page 14

Exhibit 4. DID Propensity Score Variables

BACKSTOP METHODOLOGY DOCUMENT	VARIABLES IN HKS RISK SCORE (BASED ON DCJIS DATA)	RCT COVARIATES
DEMOGRAPHICS		
age		
race/ethnicity		
RCT jurisdiction ‡		
CRIMINAL HISTORY		
# prior ~~arrests~~ arraignments*		
# prior arraignments, convictions, incarcerations	3 vars: # prior arraignments, convictions, incarceration	# prior arraignments, incarcerations (not convictions)
	4 vars: #prior arraignments with a top charge that is low, medium, medium-high, high	
	#prior arraignments w top charges of violence or drugs	
min sentence ordered for previous crimes		days of prior incarcerations
age at first arrest*	3 vars: age at first arraignment, conviction, incarceration	age at first arraignment
EMPLOYMENT		
employment history		# prior 8 quarters w > $1K earnings

*Notes: * We will use arraignments -- as in the rest of the project -- rather than arrest; arraignments initiate a DCJIS record. (I speculate that this may have been a "wordo" and was always intended to say arraignment rather than arrest.)*
‡ jurisdiction dummies can only be used for the PS for the historical group in the RCT locations, but not for the comparison jurisdictions.

6/1/24, page 15

APPENDIX 1. ASSIGNING GROUP MEMBERSHIP FOR EMPLOYMENT DATA REQUEST

The following procedure will be done separately for the RCT sample (starting with step 1), and for each of the 3 comparison groups in the DID analyses (starting with step 3).

1. For the RCT sample, the IE will divide individuals into four "treatment-status" strata defined by their **treatment referral** and **enrollment status**:
    - Referred to and enrolled[6] in Roca treatment;
    - Referred to but not enrolled in Roca treatment (i.e., noncompliance);
    - Referred to the control group and not enrolled in Roca treatment;
    - Referred to the control group but enrolled in Roca treatment (i.e., contamination).

No individuals from different treatment status strata will be grouped together.

2. For the RCT sample, each of the treatment-status strata are then divided into groups of at least 6 individuals as follows: Individuals are sorted so that adjacent individuals are as similar as possible. Then the first 6 people are grouped into group 1, the next 6 into group 2, etc. When less than 12 people remain to be sorted, they become the final group.

3. In sorting individuals, we use four variables in the following priority order:
    - **quarter of randomization**
    - **agency source type:**
        - Type 1 = Office of the Commission of Probation
        - Type 2 = Department of Corrections, Massachusetts Parole Board, Department of Youth Services, Houses of Correction
    - **geographic service area**: (Boston vs. Chelsea vs. Springfield)
    - **Age** at randomization

    We use a two-step sorting procedure. All individuals are first sorted by quarter of randomization. Then, within quarters, individuals are sorted on the three remaining characteristics: first, agency source type, then geographic service area, and then age at randomization.

    To ensure that the last person in quarter K is similar to the first person in quarter K+1, we reverse the sort order of all individuals within every other quarter. (For example, in quarter K, individuals from Agency Source Type 1 precede individuals from Agency Source Type 2; in quarter K+1, individuals in Agency Source Type 2 precede those from Agency Source Type 1. As a result, if individuals from adjacent quarters are sorted into one group, they are likely to be from the same Agency Group.)

For individuals with missing SSNs, the IE cannot include those individuals in the file sent to DUA because they are unmatchable to UI data. Instead, the IE will follow the five-step process above to assign individuals with missing SSNs to the most relevant Group ID among those *with* SSNs. The IE will then impute employment data for those missing SSNs as the average value for those with SSNs in the same Group ID.

[6] Enrollment is defined as having a Roca "Date Eligible" as of the time of OMQ file creation.

6/1/24, page 16

APPENDIX 2. EXCLUSION CRITERIA

In general, the data on possible referrals (during the RCT) and on possible comparison groups was a large superset of cases, more of which were not eligible for the program. A first task was to exclude all the ineligible cases.

In the RCT, an individual could only be referred for randomization once; similarly, in the historical period, each individual is considered for the sample only at the first eligible release to DOP.

Basic grounds for eligibility include age (17-24), the jurisdiction, and being high risk.

Then there were also some more particular bases for exclusion. The evaluator runs a series of programs on the data to exclude most of these cases and identify the eligible sample.

Risk

RCT

The samples are limited to individuals at high risk for reoffending. In the RCT, this was based on 3 different risk tools, depending on referral source and timing. DOP (including OCP referrals) used two different risk scores over the randomization period, the ORAS and the Legacy risk tools. For individuals referred by other agencies, the Harvard Kennedy School (HKS) designed a risk tool.

> "Among probationers referred from OCP, high risk is assessed based on the scores from the Legacy and ORAS risk assessment tools. Legacy scores that are 9 and lower and ORAS scores that are 21 and greater are considered high risk individuals. All other referred individuals from county and state agencies have risk scores created from a Harvard Kennedy School (HKS) code…. Scores of 135 and above are high risk." (Urban Revised Evaluation Plan, p. 4).

DID Comparison Groups

All individuals who fail to meet the high-risk definition on one of the two probation risk tools are be excluded. In the RCT, the HKS risk tool was only used for non-probation referrals – who are not part of the DID. However, as described in the section on propensity score weighting, we will include the covariates that underlying HKS's risk scores – all of which are criminal history variables (and jurisdiction) -- when creating propensity scores.

Prior Participation in Roca

The RCT evaluation excludes individuals who participated in Roca since 2011, before they were referred for randomization.

- The historical comparison group consists of individuals who were released from incarceration to DOP from 2010-2012. *We will want to identify any of them who participated in Roca in 2011 or 2012.*[7]

---

[7] As discussed above, we will *not* exclude from the historical comparison groups those individuals who then end up in the RCT, because that would bias the comparison samples.

6/1/24, page 17

Open Felonies

Individuals with open felonies at the time of entry into the samples are excluded. For the RCT samples, entry is at the time of referral for randomization; for the DID comparison samples, entry into the sample is the first eligible release to Probation during the sample period (2014 to 2018 for the contemporaneous comparison group from other jurisdictions; 2010-2012 for the program period).

Sex Offender History

Individuals with a prior conviction on a sex offense felony at the time of entry into the samples.

6/1/24, page 18

APPENDIX 3. IMPUTATION

The evaluation plan considers 2 kinds of imputation at the time of the final estimate only, for the recidivism outcome. The sections below are copied directly from the earlier evaluation plans.

1) Imputing Five-Year Bed Day Impacts for Individuals with Less than 5 Years of Observation

To better understand the table below and how it is used, I reverse engineered the table below in a memo dated 5/16/18. I found that the historical expectation for bed-days B_M in the table, following M months of observation, is:

$$B_M = (11.29776 \times M) - (0.0342169 \times M^2) - .0736995.$$

To get an "imputed" value at 60 months for a given individual, one calculates the ratio between the historic expectation at month M to the final observed bed days at month M, and then that ratio is applied to the historical expectation at 60 months.

This essentially takes the growth curve in bed days that was seen in the historic data, and applies that curve to the observed data. (If that curve is estimated as linear, the simpler curve is: $B_M = (8.83 \times M) + 37.48$, meaning a growth in expected bed days of about 9 days per month observed.)

From the earlier evaluation plans:

(Note: references to the Q14 estimates for the DOL payments are greyed, as well as the parts of the table that will not be relevant).

> The project intends to make payments based on the five-year reduction in bed days. However, at the time that measurement occurs for the final DOL payments and also at the time measurement occurs in advance of final project wrap up not all individuals will have been observed for a full five years from the date of their random assignment. Therefore, the independent evaluator will calculate "Department of Labor Total Bed-Days Avoided Per Person Served" and "Total Bed-Days Avoided Per Person Served" by adjusting upward the Actual Bed-Days Avoided Per Person Served based on the historic ratio of five-year bed-days to shorter-term bed days. The adjustment will be done as follows:
>
> Separately, as part of the evaluation process when calculating the Department of Labor Total Bed-Days Avoided Per Person Served in Q13 and the Total Bed Days Avoided Per Person Served in Q23:
>
> Count the number of Roca Youth with Months Observed [N] where N will range from 12 months and less to 60 months and will be defined as the period between the Roca Youth's random assignment date and end of the Observation Period.
>
> Calculate the Weighted Average Historical Bed Days: Multiply the number of Roca Youth with Months Observed [N] by the respective Preset Historical Bed Days specified in the table below to calculate the Total Historical Bed Days Per Month. Then sum the Total Historical Bed Days Per Month for each Observation Length to calculate

6/1/24, page 19

the Total Historical Bed Days and then divide by the number of Roca Youth to estimate the Weighted Average Historical Bed Days.

Calculate the Multiplication Factor: Divide the preset estimated 60-month historical bed days of 554.6 by the Weighted Average Historical Bed Days to calculate the Multiplication Factor.

Specifically, the Multiplication Factor will be calculated as:

$$\text{Multiplication Factor} = I_{60} \Big/ \frac{\sum_{m=12\ and\ less}^{m=60} N_m I_m}{\sum_{m=12\ and\ less}^{m=36} N_m}$$

Where I_m is the Preset Historical Bed Days for month m and N_m is the actual number of Roca Youth with Months Observed of m months.

Then Department of Labor Total Bed-Days Avoided Per Person Served will be calculated as:

Multiplication Factor X Department of Labor Actual Bed-Days Avoided Per Person Served.

Total Bed-Days Avoided Per Person Served will be calculated as:

Multiplication Factor X Actual Bed-Days Avoided Per Person Served.

In performing these calculations, the Multiplication Factor shall be rounded to four digits to the right of the decimal place.

Table of Historical Bed Days for Calculating the Multiplication Factor	
Months observed	Preset Historical Bed Days
12 and below	130.6
13	141.0
14	151.4
15	161.7
16	171.9
17	182.1
18	192.2
19	202.2
20	212.2
21	222.1
22	231.9
23	241.7
24	251.4
25	261.0
26	270.5
27	280.0
28	289.4
29	298.8
30	308.1
31	317.3
32	326.4

33	335.5
34	344.5
35	353.4
36	362.3
37	371.1
38	379.8
39	388.5
40	397.1
41	405.6
42	414.1
43	422.5
44	430.8
45	439.0
46	447.2
47	455.3
48	463.4
49	471.4
50	479.3
51	487.1
52	494.9
53	502.6
54	510.2
55	517.8
56	525.3
57	532.7
58	540.1
59	547.4
60	554.6

Example: The table below shows an illustrative example of the upward adjustment calculation. Only the Input column containing data on the number of individuals observed for each number of months will be updated as the project unfolds.

Months Observed	Input: Number of Roca Youth	Preset Historic Bed Days	Total Historic Bed Days Per Month			
	A	B	C=A*B			
12 and under	0	130.6	0.0			
13	0	141.0	0.0			
14	0	151.4	0.0			
15	10	161.7	1617.1			
16	10	171.9	1719.4			
17	10	182.1	1821.1			
18	10	192.2	1922.1			
19	10	202.2	2022.4			
20	10	212.2	2122.0			
21	10	222.1	2220.9			
22	10	231.9	2319.2			
23	10	241.7	2416.7			
24	10	251.4	2513.6			
25	10	261.0	2609.8			
26	10	270.5	2705.3			
27	10	280.0	2800.1			
28	10	289.4	2894.3			
29	10	298.8	2987.8			
30	10	308.1	3080.5			
31	20	317.3	6345.3			
32	20	326.4	6528.1			
33	20	335.5	6709.6			
34	20	344.5	6889.7			
35	20	353.4	7068.4			
36	20	362.3	7245.8			
37	20	371.1	7421.8			
38	20	379.8	7596.4			
39	20	388.5	7769.7			
40	20	397.1	7941.6			
41	20	405.6	8112.1			
42	20	414.1	8281.3			
43	20	422.5	8449.1			
44	20	430.8	8615.5			
45	20	439.0	8780.6			
46	20	447.2	8944.3			
47	20	455.3	9106.7			
48	20	463.4	9267.6			
49	20	471.4	9427.2			
50	30	479.3	14378.2			
51	30	487.1	14613.5			
52	30	494.9	14846.8			
53	30	502.6	15078.0			
54	30	510.2	15307.2			
55	30	517.8	15534.3			
56	30	525.3	15759.4			
57	30	532.7	15982.4			
58	30	540.1	16203.4			
59	30	547.4	16422.3			
60	30	554.6	16639.2			
Total Historical Bed Days (Sum Column C)						359037.9
Roca Youth (Sum Column A)						870
Weighted Average Historical Bed Days (Total Historical Bed Days/Roca Youth)						412.7
Preset Estimated 60-Month Historic Bed Days						554.6
Multiplication Factor (Preset 60-Month Historic Bed Days/Weighted Average Historic Bed Days)						1.3440

The Multiplication Factor is calculated by dividing the Preset Estimated 60-Month Historic Bed Days by the Weighted Average Historical Bed Days, which in this example is 554.6 / 412.7 = 1.3440.

6/1/24, page 22

2) Imputation of Expected Sentences for Individuals with Open Arraignments

(From the earlier evaluation plan, unchanged)

Adjusting for arraignments not yet adjudicated

For participants with an open arraignment at the final observation date, the IE will impute expected time sentenced based on the offense category (as described below) of the arraignment, with imputation conducted separately for DYS youth and Adult Probationers. For open arraignments, the IE will impute expected days of incarceration (using the median value) for that offense category multiplied by the probability that that arraignment type leads to an incarceration based on the tables below:

DYS

Description	Median Days	Average Days	N (incarcerated only)	N (all arraignments)	Fraction of arraignments resulting in Incarceration	Imputed Days
Alcohol	152	186	83	3,245	2.56%	3.9
Violent	363	533	1924	38,802	4.96%	18.0
Attempted	298	388	117	3,291	3.56%	10.6
Burglarious	182	309	780	20,252	3.85%	7.0
Drug	273	409	1993	25,210	7.91%	21.6
Disturbance	91	163	356	11,260	3.16%	2.9
Firearm	547	838	751	13,305	5.64%	30.9
Larcenous	182	295	841	17,324	4.85%	8.8
Deceitful	182	280	660	11,403	5.79%	10.5
Motor Vehicle	90	184	1231	16,980	7.25%	6.5
Other	182	388	145	4,547	3.19%	5.8
Obstruction of Justice	182	244	404	5,712	7.07%	12.9
Property	182	283	479	16,517	2.90%	5.3
Robbery	912	1084	244	6,523	3.74%	34.1
Sex or Indecent	273	418	287	4,859	5.91%	16.1
Threat of	182	287	291	6,528	4.46%	8.1

Adult Probation

Description	Median Days	Average Days	N (incarcerated only)	N (all arraignments)	Fraction of arraignments resulting in Incarceration	Imputed Days
Alcohol	90	150	159	24,195	0.66%	0.6

Violent	257	426	4546	143,783	3.16%	8.1
Attempted	273	424	290	13,115	2.21%	6.0
Burglarious	182	308	1804	78,271	2.30%	4.2
Drug	182	384	5446	129,890	4.19%	7.6
Disturbance	90	162	894	46,117	1.94%	1.7
Firearm	547	790	1386	38,431	3.61%	19.7
Larcenous	182	282	2304	74,383	3.10%	5.6
Deceitful	182	293	1562	48,425	3.23%	5.9
Motor Vehicle	90	164	3941	109,419	3.60%	3.2
Other	182	351	448	23,904	1.87%	3.4
Obstruction of Justice	182	249	917	20,643	4.44%	8.1
Property	182	262	1111	70,038	1.59%	2.9
Robbery	730	973	539	17,080	3.16%	23.0
Sex or Indecent	182	355	778	19,520	3.99%	7.3
Threat of	182	276	748	24,843	3.01%	5.5

Notes: Tables include data from DYS males aged 17-24 and "high-risk" probationers (adult and juvenile) aged 17-24. Omitting Juvenile Probationers does not meaningfully affect results.

The table below shows the offense category, and the CORI codes that feed into each offense type:

Offense Type	CORI codes
Violent	A&B, ASLT, MURD, MANS, HOMICIDE, MAYHEM
Burglarious	B&E, TRES, HOME INVASION, BURG, PBT, ENT WO BRK
Deceitful	RSG, CNTRFT, IDENT, TRUE NM, FRD, UTTER, FORG, F&U, CRDT CARD, EXTORT, DISGUISE, DFRD INNKPR, IMPERS, ID VIOL
Attempted	CONSP, ATT, ACC AFT, ACC BEF
Drug	CSA
Robbery	ROB, CARJACK, KDNP
Larcenous	LAR, SHOPLIFT, THEFT
Disturbance	DP, DIS PERS, DIS COND, DPA, DSA, DISORD, AFFRAY, FAIL TO DISPERSE
Threat of	THREAT, INTIM, STEAL

Firearm	FIR , FIR$, AMMO DWC, AIR RIFLE, POSS DW, POSS ELECTR WEAP, EXPL, BMB HX, POSS MACE
Property	PROP, VAND, BRN, GLS BRK, ARSON, INJ BLDG
Motor vehicle	11[0-9], MV, ^MV
Sex or indecent	RAPE, IND, SEX, PORN, AB PREV, O&G
Alcohol	LIQ, ALC, POSS OPEN CONT, DRINK PUB
Obstruction of justice	RESIST ARST, FL OBEY PO, FLS ADDR HIND PO, REF ID PROC, FL APR PER RECOG, OBSTR JUST, PO INTF, FLS STMNT, ESCP, OFF REF AID

APPENDIX 4. REGRESSION SPECIFICATION

Recidivism Models (written in Stata)

ITT: regress RecidDays i.Quarter i.Agency##i.City ///
PFSAge Age1st_arr Arr_pre_total Inc_pre_total Days PreEmpl TxReferred [aweight=WeightNew], robust[8]

IVE: ivregress 2sls RecidDays i.Quarter i.Agency##i.City ///
PFSAge Age1st_arr Arr_pre_total Inc_pre_total Days PreEmpl (TxEnrolled=TxReferred) [aweight=WeightNew], vce(robust) first[9]

Employment Models (written in Stata)

The employment models are estimated at the Group ID level and are weighted using the sum of the weights for all individuals within a group (which may include individuals without an SSN). Although the employment models are estimated at the Group ID level, the outcome will represent the per-person average.

ITT: regress meanPostEmpl ///
meanQ2 meanQ3 meanQ4 meanQ5 meanQ6 meanQ7 meanQ8[10] meanAgency1City1 ///
meanAgency2City1 meanAgency1City2 meanAgency2City2 meanAgency1City3 ///
meanPFSAge meanAge1st_arr meanArr_pre_total meanInc_pre_total meanDays meanPreEmpl
TxReferred [aweight=GroupWeight], robust[11]

IVE: ivregress 2sls meanPostEmpl ///
meanQ2 meanQ3 meanQ4 meanQ5 meanQ6 meanQ7 meanQ8 meanAgency1City1 ///
meanAgency2City1 meanAgency1City2 meanAgency2City2 meanAgency1City3 ///
meanPFSAge meanAge1st_arr meanArr_pre_total meanInc_pre_total meanDays meanPreEmpl
(TxEnrolled=TxReferred) [aweight=GroupWeight], vce(robust) first[12]

---

[8] Specifies calculation of robust standard errors.
[9] Specifies that output shows the first and second stages of the two-stage least squares regression.
[10] For categorical variables, the last category (in this case, meanQ9) will be omitted as a reference.
[11] Specifies calculation of robust standard errors.
[12] Specifies that output shows the first and second stages of the two-stage least squares regression.

APPENDIX 5:

BACKSTOP METHODOLOGY

Original Plan Edited by the Urban Institute

May 2015

Problem #1: Lack of Precision

The problem is that the study may lack sufficient precision based on treatment group sample size and rates of compliance with group assignment. This may occur for a variety of reasons. For example, the sample size could be insufficient because there are too few Roca Assigned Youth on the Roca List, because there are too few Conforming Referrals, or because there are too few Enrollments from Referrals. The contract attempts to safeguard against such risks to precision by stating that the RCT must include at least 765 Enrollments into Roca from Referrals by the end of Q14 to be deemed sufficient for a reliable estimate of Roca's effectiveness. If the treatment group size fails to meet this minimum (i.e., is less than 765 individuals), we do not propose augmenting the treatment group by including other individuals served by Roca because doing so would introduce too much bias into the estimate.

While there is no viable method that improves the precision of the estimate of the treatment group outcomes, a backstop methodology can marginally increase the precision of the estimate of the *differences* between the treatment and control groups. This methodology, documented herein, combines the RCT with a quasi-experimental estimate of the treatment effect. Thus, if the treatment group fails to meet the required threshold of 765 individuals, then this backstop will be employed.

Precision also depends on the size of the control group. The parties have agreed that if the control group (i.e. all individuals assigned to the control group) has less than 500 individuals (calculated as 1/3 of the total expected referrals) then the same quasi-experimental methodology described herein will be employed.

Problem #2: Treatment and Control Crossover

A related problem for the design is if too much crossover between treatment and control categories occurs. That is, if too many of the individuals assigned to the control category enroll in and receive services from Roca and/or if too many of the individuals assigned to the treatment category *do not* receive Roca services. To safeguard against such a risk to precision, the contract states that the difference in the fraction of Roca Assigned Youth and Control Youth that enroll in Roca must be greater than or equal to .30 for the RCT to have sufficient accuracy. If this minimum difference in the fractions served (the "compliance fraction") is not met, the quasi-experimental backstop methodology described herein will be employed in conjunction with the RCT to obtain a marginally more precise estimate of the differences between groups by increasing the size of the control group.

**PROPOSED FRAMEWORK FOR
BACKSTOP EVALUATION METHODOLOGY**

This document outlines the framework for the backstop evaluation methodology.

As agreed to by the parties, the RCT approach will have sufficient accuracy to estimate Roca's effectiveness if there are at least 765 Enrollments from Referrals, the control group is has at least 500 individuals, and the difference in the compliance fraction of Roca Assigned Youth and Control Youth that enroll in Roca is greater than or equal to .30. Otherwise the sample sizes and/or difference in the "treatment dosage" received by the treatment and control groups will not be large enough for the RCT approach alone to yield a sufficiently precise estimate of the impact of being enrolled in Roca.

As described previously, if the required Enrollments from Referrals, control group size, and difference in compliance fraction are not achieved, a secondary, quasi-experimental approach will be employed in order to obtain a sufficiently precise estimate. This backstop methodology is inferior to a successful RCT since differences in outcomes between the referral and non-experimental comparison populations could be due to factors other than the experimental intervention—for example an improving economy or changes in policing or sentencing patterns. Nonetheless, because payments will be flowing based on the estimated outcomes, it is important that the evaluation methodology be robust enough to produce reliable estimates even if the RCT methodology is not successful in producing a sufficiently large group of Enrollments from Referrals, control group, or difference in compliance fractions.

Statistical simulations by the Harvard SIB lab suggest that as long as there are at least 765 Enrollments from Referrals, 500 individuals in the control group, and a difference between P_T and P_C of at least 0.30, the RCT will have reasonable statistical power and there will be no need for a backstop. These simulations show that when these requirements are not met, the RCT estimates are still valuable, but that the overall precision of the estimates can be improved by combining the RCT estimate with a quasi-experimental approach.

If there are not at least 765 Enrollments from Referrals, or 500 individuals in the control group, or the difference between P_T and P_C falls below 0.30, then we will remedy this imprecision as follows: First, we will produce a "difference in difference" estimate of Roca's effect using the methodology detailed below. Second, we will combine this estimate with the RCT IV estimate by assigning a 0.20-weight to the backstop value and a 0.80-weight to the RCT value.[13]

---

[13] This weighting ratio is based on statistical simulations conducted by the Harvard SIB lab using rough approximations of three possible sources of imprecision. The first source of imprecision is that resulting from unequal sample sizes for the RCT and backstop analyses—which is assumed to be zero based on examination of the current project data. The second source of imprecision is that resulting from the non-experimental natures of the backstop methodologies—which is assumed to be high, given the possible confounding explanations for any observed treatment-control differences. The third source of imprecision

Differences in Differences
The backstop methodology is based on a "difference in differences" (DID) methodology. This methodology will compare the change in outcomes before and after the intervention was introduced among eligible individuals from the PFS Project Areas (i.e., Boston, Chelsea, and Springfield areas) with the change in outcomes over the same time period in two different comparison populations. The two comparison populations are:

1. Same age and risk, different geography. Individuals who otherwise meet study eligibility criteria and reside in the following cities where Roca does not operate will be included in this first comparison population.

- *Brockton Area:* Brockton
- *Fall River Area:* Fall River; New Bedford
- *Lawrence Area:* Methuen; Haverhill; Lawrence; Lowell
- *Worcester Area:* Worcester

2. Same risk and geography, older ages. Individuals in the PFS Project Areas who meet all eligibility criteria except for age and whose age is slightly above the level for eligibility[14] in the PFS project will be included in this second comparison population.

This differences-in-differences approach will compare these populations to those individuals from the RCT who are referred to Roca. These individuals are a random subset of all eligible individuals in the PFS Project Areas after the introduction of the intervention.

The backstop methodology will construct estimates using each of these two approaches and the backstop estimate will be the average of the two (i.e., each approach will receive a 0.50-weight in determining the backstop methodology's estimate of Roca's effectiveness).

<u>Estimating the Propensity Score</u>
The backstop methodology will utilize a propensity score methodology to ensure that the individuals in the two groups being compared in each comparison are as similar as possible in their distributions of observable characteristics. In particular, the DID procedure will weight each individual by a function of his *propensity score*.

---

is that resulting from the fact that the treatment group in backstop calculations overlaps with that in RCT calculations.

[14] Starting with men ages 25 and increasing as necessary to reach a sample size roughly equivalent to that of the RCT.

The *propensity score* is the estimated probability that a young man is in the PFS Project Sample rather than a Comparison Sample, based on individual-level characteristics. The IE will estimate the propensity score via the following logistic regression:

$$PFS_i = g\left(\alpha + \sum_{k=1}^{K} \beta_k X_{ik}\right)$$

where PFS_i is a binary indicator for whether individual i is in the PFS Project Sample; α is the overall intercept; X_{ik} is the *k*th covariate for individual *i*, with associated coefficient β_k; and $g(\cdot)$ is the logit link function.

The potential covariates to be used include:

- Age
- Race/ethnicity
- Risk score
- Number of previous arrests
- Age of first arrest
- Arraignment count (defined as pre-random assignment for treatment and control, and a calculation of that timeframe for the new comparison)
- Conviction count (defined as pre-random assignment for treatment and control, and a calculation of that timeframe for the new comparison)
- Incarceration count (defined as pre-random assignment for treatment and control, and a calculation of that timeframe for the new comparison)
- Minimum sentence ordered for previous crimes (defined as pre-random assignment for treatment and control, and a calculation of that timeframe for the new comparison)
- Educational attainment
- Employment history

<u>Trimming and Propensity Score Balance</u>
The propensity scores will be checked for balance and overlap. If the propensity scores generate extreme weights, these weights will be trimmed.

<u>Propensity Score Weighting</u>

The weights for this part of the DID analysis are very similar to the RCT weights, except with estimated quantities rather than using the observed proportion of individuals. The IE will calculate estimation weights using the following[15]:

- The weight for each individual in the PFS Project Sample will be 1.
- The weight for each individual, *i*, in the Comparison Samples will be based on their propensity scores (PS):

$$W_i = \frac{1 - \widehat{PS}_i}{\widehat{PS}_i}$$

Where $\widehat{PS}_i$ is the estimated propensity score for individual *i*.

Ensure similar observation windows and risk levels

The analyses will need to be set up to ensure that the observation window for the project areas and comparison areas are similar. The analysis will also need to ensure that individuals in earlier time periods have similar risk levels to individuals in the experimental time period; doing so is complicated by the fact that the state uses a different risk scores today than it did in the past.

Calculate base DID ITT effect

Analogously to the RCT, the DID ITT estimate is defined as the difference between the average outcomes for those individuals in a Project Sample (PS) and those in a Comparison Sample (CS), regardless of whether the young men enroll in Roca:

$$\widehat{ITT}^{DD}_{base} = \hat{Y}^{PS} - \hat{Y}^{CS}$$

where $\hat{Y}^{PS}$ and $\hat{Y}^{CS}$ are estimated by applying the weights described in the previous subsection to the observed outcomes, Y_i:

$$\hat{Y}^{PS} = \frac{\sum_{i=1}^{N_{PS}} Y_i^{PS} W_i^{PS}}{\sum_{i=1}^{N_{PS}} W_i^{PS}} = \frac{\sum_{i=1}^{N_{PS}} Y_i^{PS}}{N_{PS}}$$

$$\hat{Y}^{CS} = \frac{\sum_{j=1}^{N_{CS}} Y_j^{CS} W_j^{CS}}{\sum_{j=1}^{N_{CS}} W_j^{CS}}$$

N_{PS} and N_{CS} are the number of individuals in the Project and Comparison Areas, respectively; Y_i^{PS} is the outcome (number of bed days or employment status) for each individual (indexed by i) in the Project Areas and Y_j^{CS} is the same for Comparison Areas (indexed by j); W_i^{PS} is the weight for

[15] Abadie, A. 2005, "Semiparametric Difference-in-Differences Estimators," Review of Economic Studies, vol. 72(1), 7-9.

each individual in the Project Areas (all set to 1) and W_j^{CS} is the weight for each individual in the Comparison Areas. Note that both $\sum_{i=1}^{N_{PS}} W_i$ and $\sum_{j=1}^{N_{CS}} W_j$ equal N_{PS}.

Estimating the overall DID ITT effect
Each of the DID estimates will be calculated by comparing the change over time in outcomes for the project sample to the change in outcomes over time for one of the comparison samples.

$$\widehat{ITT}_{DD} = \left(\hat{Y}^{PS} - \hat{Y}_{pre}^{PS}\right) - \left(\hat{Y}^{CS} - \hat{Y}_{pre}^{CS}\right)$$

Adjusting the Difference-in-Difference Estimate for Non-Compliance
As with the RCT, it is necessary to adjust the ITT estimate to calculate Roca's effect on a per-person served basis. In this case, the adjustment assumes that no individuals in the Comparison Samples received Roca services. In that case, the IE will calculate the IV estimate as follows:

$$\widehat{IV}^{DD} = \frac{\widehat{ITT}^{DD}}{\hat{p}^{PS}}$$

where $\hat{p}^{PS}$ is the proportion of individuals in the RCT Roca referral group who enrolled in Roca. Implicitly, this calculation uses the same weights as the overall ITT calculation; however, since the weights equal 1 for all individuals in the RCT Roca referral group, this is a simple proportion.

Finally, in order to combine this estimate with the RCT IV estimate, we also assume that the impact of Roca on those individuals who actually enroll in the program is the same regardless of whether they are on the Roca List.

ATTACHMENT B: ABT ANALYSIS

Reincarceration among Roca participants in Massachusetts

Authors: Shelby Hickman, Claudia Masters, Nikitha Reddy

Introduction

Roca serves young men at the greatest risk of involvement in violent crime, including those who have served time in jail or prison for violent offenses. In this brief, we describe reincarceration rates for Roca Massachusetts participants one, two, and three years after they join Roca.

Key Takeaways

We find that a relatively small proportion of Roca participants are reincarcerated.

- **Roca's 2017 cohort shows lower reincarceration rates than the state average** at one, two, and three years follow-up, **even for those with a history of violent offenses**.
- Over time, the gap widens, with **Massachusetts' three-year rate 30% higher than Roca's.**

Methods

Data. These analyses draw on Roca's program data and Criminal Offender Record Information provided by the Department of Criminal Justice Information Services in Massachusetts.

Sample. The sample includes 18- to 24-year-old male Roca participants at the Boston, Chelsea, Lynn, Holyoke, and Springfield sites with a history of incarceration in Massachusetts prior to joining Roca. We describe reincarceration rates for participants with a history of any incarceration prior to joining Roca programming, as well as those with a history of incarceration for a *violent offense* who are reincarcerated for a *violent offense*.

Our sample is designed to match the recidivism cohort structure used by Massachusetts for its recidivism analysis. Sample sizes for each annual cohort are described in Table 1.

Table 1. Sample size by annual cohort

Cohort	2013	2014	2015	2016	2017
Roca participants	94	249	174	222	223
Roca participants with a history of incarceration	44	165	128	136	151
Roca participants with a history of incarceration for a violent offense	31	109	86	81	92
Massachusetts cohort, total releases[a]	N/A	N/A	N/A	N/A	1,104

a: Source: https://www.mass.gov/info-details/cross-tracking-system-recidivism-query-model

A Roca cohort includes male participants 18-24 who joined Roca in a given calendar year. The state cohort includes all men ages 18-24 who were released from a House of Corrections, Department of Corrections facility, or jail in a calendar year.

Comparison of Roca reincarceration versus Massachusetts reincarceration rates.

We compare the rates of reincarceration for Roca participants to rates of reincarceration for the Massachusetts sample in the 2017 cohort. This is the best reference group we can obtain using publicly available reports. We use 2017 because this is the most recent year that state data are available where at least two subsequent years are not affected by the COVID-19 pandemic. During the COVID-19 pandemic, changes to arrest, release, transfer, and court policies and procedures are associated with unusual trends for incarceration and reincarceration. Any comparisons we make are meant only to provide a reference point for the Roca calculations. We cannot make any inferences about statistical differences between the Roca sample and Massachusetts sample for three key reasons:

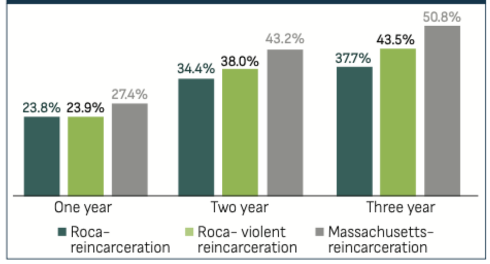
abtglobal.com

1. To conduct a statistical test of differences for two proportions, the two samples must be independent of each other. In this case, members of the Roca sample who are reincarcerated are also included in the Massachusetts sample, which means the samples are not independent.
2. The Massachusetts reincarceration rate counts each *release* during the year, making it possible for one person to be included multiple times if they are incarcerated more than once in the same calendar year. Our calculations for Roca participants include each *individual* only one time per calendar year. This means that if someone is reincarcerated twice in one year, it counts one time in our Roca estimate, but twice in the Massachusetts estimate. Although it is uncommon for a participant to enroll with Roca more than one time per year, this methodological difference can elevate the rate of reincarceration estimated for the state cohort when compared to rates we report for the Roca cohort.
3. The "clock" for our reincarceration calculations for Roca participants begins the day they join Roca. The state, by contrast, begins its "clock" the day a person is released from jail or prison. Because the starting point for each group is different, the Roca estimates at one, two, and three years after joining Roca, can actually be more than one, two, and three years after release, depending on the amount of time between release and joining Roca. This means Roca estimates might be biased toward higher levels of incarceration than the state comparison for reincarceration.

Results and Discussion

Although Roca serves only the most high-risk individuals, its participants have lower rates of reincarceration than are seen for Massachusetts in the same cohort (2017). As shown in Figure 1, for the 2017 cohort, we see that Roca participants consistently have lower rates of reincarceration than the state sample at one, two, and three years. Among Roca participants at highest risk for reincarceration — those who have previously been incarcerated for a violent offense — rates of reincarceration for a violent offense are also lower than the

Exhibit 1. One-, two-, and three- year reincarceration rates for Roca and Massacusetts 2017 cohorts



Massachusetts total reincarceration rates. Rates for each Roca annual cohort 2013-2017 are summarized in Table 2, below.

These trends show promise for Roca's program. Further, the difference between Roca and the state reincarceration rates for the 2017 cohort grows over time and is greatest at three years, where the Massachusetts rate is 30% higher than Roca's. This suggests that longer engagement with Roca could have added marginal benefit for participants. Extended engagement is central to Roca's program model, which lasts four years. These trends show that, indeed, longer engagement seems to have a payoff. Given that Roca engages participants during the period of the life course when offending rates are highest (ages 18-24), it is possible that any impacts of Roca programming on incarceration are sustained as participants age out of this high-risk period.

The analyses presented in this brief, however, do not isolate the effect of Roca's programing. We present reincarceration rates without controlling for factors known to influence likelihood of offending and incarceration. Factors such as services received in jail, social and family supports, employment, and conditions of parole or probation can all affect likelihood of incarceration and would provide a more nuanced understanding of recidivism among Roca's participants. Overall, even without the benefit of a direct comparison group that would tell us what incarceration would have been if these participants did not enroll in Roca, only a small proportion of Roca participants are reincarcerated.

Table 2. Roca reincarceration rates at one, two, and three years by cohort

Follow up period	One Year					Two Years					Three Years				
Cohort	2013	2014	2015	2016	2017	2013	2014	2015	2016	2017	2013	2014	2015	2016	2017
Roca reincarceration	16%	29%	25%	27%	24%	39%	40%	38%	38%	34%	48%	47%	47%	50%	38%
Roca violent reincarceration	19%	29%	27%	32%	24%	45%	42%	37%	38%	38%	48%	51%	47%	47%	44%

Abt Global is a consulting and research firm that has combined data and bold thinking to improve the quality of people's lives since 1965. For nearly 60 years, we have partnered with clients and communities to advance equity and innovation—from creating scalable digital solutions and combatting infectious diseases, to mitigating climate change and evaluating programs for measurable social impact. Our global workforce crosses geographies, methods, and disciplines to deliver tailored solutions in health, environmental and social policy, technology, and international development.

Connect With an Abt Expert

abtglobal.com

Printed on 100% Recycled Content Paper | 2024

ATTACHMENT C: CONTRACT CHANGES

The table below details each of the contract negotiations and changes in the changes in the assumptions applied to the project over time.

ATTACHMENT C

	Proposed Changes to the Contract and/or Project	# Randomized	# Served	% of Treatment Group to Be Treated	Expected Recidivism Rates	Expected Days of Incarceration
Initial Assumption	N/A	Treatment: 1,514 Control: 1,514 Total: 3,028	1060	80%	1 yr: 31.7% 2 yr: 44.9% 3 yr: 51.8% 4 yr: 56.5% 5 yr.: 59.5%	Sentenced: 860 Total: 520

ATTACHMENT C

	Proposed Changes to the Contract and/or Project	# Randomized	# Served	% of Treatment Group to Be Treated	Expected Recidivism Rates	Expected Days of Incarceration
Initial Contract (1/7/14)	Initial Contract – served 21 cities across Commonwealth	Treatment: 1,327 Control: 1,327 Total 2,654	929 (929 from state referrals)	80%	1 yr: 31.9% 2 yr: 45.4% 3 yr: 52.4% 4 yr: 57.0% 5 yr: 59.9%	Sentenced: 930 Total: 560

ATTACHMENT C

	Proposed Changes to the Contract and/or Project	# Randomized	# Served	% of Treatment Group to Be Treated	Expected Recidivism Rates	Expected Days of Incarceration
Side Letter (5/20/14)	Payment for shortage of participants in first month of contract $750,000			80%		
Side Letter (6/30/14)	Roca allowed to Self-recruit 54 young men in Q4			80%		

ATTACHMENT C

	Proposed Changes to the Contract and/or Project	# Randomized	# Served	% of Treatment Group to Be Treated	Expected Recidivism Rates	Expected Days of Incarceration
Amended Contract 1 (12/16/14)	Added in referrals from four houses of corrections, Department of Corrections and Parole. Added in ability for Roca to self-recruit participants to maintain financial viability in light of reduced enrollment.	Treatment: 1,272 Control: 1,272 Total 2,548	929 (584 from state referrals, 345 from self-recruits	53%		
Side Letter (5/18/15)	Agreement to extend deadline for Backstop Evaluation					
Side Letter (7/26/16)	Extension of timing for renegotiation to keep Commonwealth out of breach for low referrals					
Side Letter (9/30/16)	• Extension of timing for renegotiation to keep Commonwealth out of breach for low referrals					

	Proposed Changes to the Contract and/or Project	# Randomized	# Served	% of Treatment Group to Be Treated	Expected Recidivism Rates	Expected Days of Incarceration
Amended Contract 2 (11/1/16)	• Automatic effectiveness of wind-down, with opportunity to continue project if Lenders so choose • Conforming referral floor adjusted to conservative quarterly targets (20/29). 525 conforming referrals needed over enrollment period for sufficient statistical power • New enrollment target of 1036 youth. 117 new youth funded by Laura and John Arnold Foundation • Conforming referral shortfall payment mechanism implemented • Job Readiness calculation payment method revised (measurement period 10 Qs, pro-rated, and projection) • Adjustments to funding plan based on additional grant funding ($1.67M from Laura and John Arnold Foundation) and lender draws • Revised Junior and Senior Loan Agreements • Reduction in number of target slots filled per quarter by Roca (drive funding draws)	Treatment: 1,349 (632 conforming) Control: 1,349 Total: 2,698	1,036 (530 from state referrals, 506 from self-recruits)	43% (48% conforming with Roca accepting 88% of conforming)	1 yr: 31.9% 2 yr: 45.4% 3 yr: 52.4% 4 yr: 57.0% 5 yr: 59.9%	Sentenced: 930 Total: 560
Amended Contract # 3 (1/1/2020)	• Extended the observation period of the Evaluation by 15 quarters to account for the "low and slow" number of referrals. • Commonwealth agreed to pay $4M in Retention and Completion Payments to Roca and $.5 in expenses related to the extension. • Postponed final report from 12/31/19 to 3/31/24.					

The result of these renegotiations, all designed to continue the project and prevent a breach of contract or a fight over the same, was a reduction in the size of the overall project. The following illustrates actual randomization results compared to both the original contract and version signed on November 1, 2016, as of March 6, 2019:

ATTACHMENT C

- Young people randomized – 38.9% less than originally planned, 31.13% less than 2016 contract.
- Young people randomized to the control group – 57.5% less than originally planned, 54.1% less than 2016 contract.
- Young people randomized to the treatment group – 20.2% less than originally planned, 10.4% less than 2016 contract.
- Conforming referrals sent to treatment – 65.5% less than originally planned, 17.2% less than 2016 contract.
- Referrals who received treatment – 50.2% less than originally planned, 12.6% less than 2016 contract.

ATTACHMENT D: ACTUAL NUMBER OF CONFORMING REFERRALS RECEIVED

The final contract set a contractual target (632) for conforming referrals and a contractual floor (525). The Commonwealth was required to pay a fee based on the number of conforming referrals below the target but above the floor. If the Commonwealth fell below 525 conforming referrals, they would be in breach of contract. The following table details the referrals to treatment and control, conforming referrals, and number of young people treated in each quarter of the project.

Referral Quarter	Control	Treatment	Total	% Control	% Treatment	Conforming Referrals	% Conforming Referrals	Treated	% of Treatment Treated	% Conform Referrals Treated	Self - Recruits
Q2: Jan-Mar 2014	41	143	184	22.3%	77.7%	61	42.7%	58	40.6%	95.1%	36
Q3: Apr - Jun 2014	23	77	100	23.0%	77.0%	34	44.2%	32	41.6%	94.1%	11
Q4: Jul - Sep 2014	41	122	163	25.2%	74.8%	73	59.8%	61	50.0%	83.6%	49
Q5: Oct - Dec 2014	36	102	138	26.1%	73.9%	52	51.0%	40	39.2%	76.9%	26
Q6: Jan - Mar 2015	24	79	103	23.3%	76.7%	35	44.3%	29	36.7%	82.9%	24
Q7: Apr - Jun 2015	28	77	105	26.7%	73.3%	40	51.9%	36	46.8%	90.0%	19
Q8: Jul - Sep 2015	19	57	76	25.0%	75.0%	27	47.4%	24	42.1%	88.9%	13
Q9: Oct - Dec 2015	35	92	127	27.6%	72.4%	32	34.8%	26	28.3%	81.3%	8
Q10: Jan - Mar 2016	45	57	102	44.1%	55.9%	27	47.4%	20	35.1%	74.1%	1
Q11: Apr - Jun 2016	31	43	74	41.9%	58.1%	22	51.2%	20	46.5%	90.9%	19
Q12: Jul - Sep 2016	33	56	88	37.5%	63.6%	28	50.0%	20	35.7%	71.4%	33
Q13: Oct - Dec 2016	30	35	65	46.2%	53.8%	23	65.7%	19	54.3%	82.6%	31
Q14: Jan - Mar 2017	54	61	115	47.0%	53.0%	24	39.3%	22	36.1%	91.7%	44
Q15: Apr - Jun 2017	29	36	64	45.3%	56.3%	16	44.4%	15	41.7%	93.8%	35
Q16: Jul - Sep 2017	39	40	78	50.0%	51.3%	13	32.5%	10	25.0%	76.9%	35
Q17: Oct - Dec 2017	23	26	49	46.9%	53.1%	14	53.8%	13	50.0%	92.9%	27
Q18: Jan - Mar 2018	37	33	68	54.4%	48.5%	7	21.2%	4	12.1%	57.1%	23
Q19: Apr - Jun 2018	27	28	55	49.1%	50.9%	13	46.4%	10	35.7%	76.9%	26
Q20: Jul - Sep 2018	21	22	43	48.8%	51.2%	10	45.5%	8	36.4%	80.0%	38
Q21: Oct - Dec 2018	28	27	55	50.9%	49.1%	12	44.4%	7	25.9%	58.3%	60
Summary:	**644**	**1213**	**1852**	**34.8%**	**65.5%**	**563**	**46.4%**	**474**	**39.1%**	**84.2%**	**558**