



**Process Study of the  
U.S. Department of  
Labor’s “Pay for  
Success” Pilots in  
Two States:  
Final Report**

June 30, 2020

*Prepared for:*

U.S. Department of Labor  
Employment and Training  
Administration  
200 Constitution Avenue, N.W.  
Washington DC 20210

*Submitted by:*

*Eliza Kean  
Deena Schwartz  
Bethany Boland  
Maureen Sarna  
Karin Martinson*

**Abt Associates**

6130 Executive Blvd.  
Rockville, MD 20852

**Acknowledgements .....iii**

**Executive Summary ..... 1**

**1. An Introduction to Pay for Success and DOL’s PFS Grants ..... 9**

1.1 The Pay for Success Approach ..... 10

1.2 The Approach to PFS Used by DOL ..... 11

1.3 The Pay for Success Field since the Launch of the DOL Pilots ..... 12

1.3.1 Results from Other PFS Projects ..... 13

1.3.2 Current State of the PFS Field ..... 15

1.4 Process Study Research Questions and Data Sources ..... 16

1.5 Structure of the Report ..... 17

**2. The Design and Implementation of the PFS Pilots ..... 19**

2.1 Launch and Early Operation of the PFS Pilots ..... 19

2.1.1 PFS Model Development Prior to SGA Release ..... 19

2.1.2 Release of the DOL SGA, Two-Phase Implementation, and Selection of Partners ..... 21

2.1.3 Development of Contracts to Manage Pilots ..... 24

2.1.4 Raising Capital and Identifying Investors ..... 25

2.2 Developing the PFS Pilots’ Evaluation Designs ..... 30

2.2.1 Specifying Evaluation Designs, Sample Sizes, and Data Collection ..... 31

2.3 Summary ..... 33

**3. PFS Pilot Interventions ..... 35**

3.1 The Intervention in Massachusetts: Roca and its Service Model ..... 35

3.1.1 Services Provided by the Roca Program ..... 36

3.1.2 Changes in the Roca Program under the PFS Pilot ..... 38

3.2 PFS Pilot Intervention in New York ..... 39

3.2.1 CEO Program Model ..... 40

3.2.2 Changes in the CEO Program for the PFS Pilot ..... 41

3.3 Summary ..... 43

**4. Random Assignment and Enrollment Challenges in the PFS Pilots ..... 44**

4.1 Implementation of Random Assignment in the Massachusetts Pilot ..... 44

4.1.1 Overview of the Random Assignment Procedures ..... 44

4.1.2 Challenges in Identification and Referral of Treatment Group Members ..... 45

4.1.3 Efforts to Address Low Enrollment in the Massachusetts Pilot ..... 47

4.2 Implementation of Random Assignment in the New York PFS Pilot ..... 49

4.2.1 Random Assignment Procedures in the NY Pilot ..... 49

---

4.2.2 Challenges in Random Assignment and Enrollment ..... 51

4.3 Summary..... 53

**5. Evaluation Findings and Payment Results..... 54**

5.1 Measures of the Intervention’s Effects ..... 54

5.2 Evaluation Results and Validation ..... 56

5.2.1 Evaluation Results and Validator Findings in Massachusetts ..... 56

5.2.2 Evaluation Results and Validator Findings in New York ..... 58

5.3 Evaluation-Validation Reports to DOL and Payment Decisions..... 59

5.3.1 Massachusetts ..... 60

5.3.2 New York ..... 60

5.4 Summary..... 61

**6. Key Findings and Suggestions for Future PFS Pilots ..... 62**

    Concluding Thoughts ..... 65

**References ..... 66**

**Appendix A. Pilot Interviewees ..... 69**

**Appendix B. Basic Concepts in Evaluation ..... 70**

**Appendix C. Recommendations for PFS Validator Reports ..... 73**

### Acknowledgements

The efforts of several individuals have been indispensable in the development of the Process Study of Department of Labor’s Pay for Success (PFS) Pilots Final Report. We gratefully acknowledge financial support and technical guidance from the U.S. Department of Labor’s Employment and Training Administration. The Contracting Officer’s Representative, Charlotte Schifferes, played a critical role in guiding the work of this study. We also acknowledge and appreciate the input from the two PFS Pilots and their many partners.

At Abt Associates, Amanda Parsad, Lisa Setrakian, and Alan Werner also contributed to the development of the report..

### Disclaimer

*This report was prepared for the U.S. Department of Labor (DOL), Employment and Training Administration under contract number Contract # DOLQ121A21884, Task Order: DOL-ETA-15-U-00030. The views expressed are those of the authors and should not be attributed to DOL, nor does mention of trade names, commercial products, or organizations imply endorsement of same by the U.S. Government.*

## Executive Summary

### Introduction

In September 2013, the U.S. Department of Labor (DOL) awarded two four-year grants to operate pilot projects to test the use of a Pay for Success (PFS) approach. Under PFS, financing for the initial costs of services comes from investors who are repaid, possibly with a substantial return, only when specific, pre-determined results are achieved and verified.

The PFS model was envisioned as an approach for potentially helping payors – often state and local governments – fund services and produce long-term benefits, such as reduced programmatic costs, more efficient spending, and better outcomes for people and communities. PFS potentially offers to governments and other payors a way to finance services that might not otherwise be feasible due to limited budgets. Beyond this primary goal, PFS initiatives are thought to have the potential to influence how payors will operate services in the future, by possibly creating incentives for: innovation, effective preventive or proactive services, flexible service delivery, and adaptable management structures.

The DOL pilots were the first effort by a federal agency to support the PFS model. However, the first project testing Social Impact Bonds—a model with similar principles as PFS—was launched in September 2010 at the Peterborough Prison in United Kingdom. Since that time, many PFS projects have been launched in the US and in other countries, and the PFS field has evolved considerably to cover a range of policy areas, including early childhood, employment, criminal justice and homelessness.

These projects have yielded information for program administrators, researchers, policymakers and other interested stakeholders that has shaped thinking on PFS. While some PFS projects made payments to investors, others did not achieve their target outcomes. When payments have not been made, projects may still be viewed by some observers as successful implementation of the PFS model: the government was able to take a risk on providing promising services without bearing the cost of services that ultimately were not found to achieve pre-determined results. At the same time, projects that have met goals for improved outcomes and have had payments made have been subject to some scrutiny due to a lack of confidence in evaluation findings or the significant payment amount made to investors.

The DOL grants were awarded to two state workforce agencies: the Massachusetts Executive Office of Labor and Workforce Development (EOLWD) and the New York State Department of Labor (NYSDOL) and were for approximately \$12 million each. Both states also committed state funds to continue the pilots beyond the DOL four-year grant period into a second phase. Both pilots focus on improving employment outcomes and reducing recidivism among individuals newly released from incarceration. The Massachusetts pilot targeted young male parolees and probationers, while the New York pilot focused on high-risk adult parolees.

Funds for the PFS pilots were drawn from an appropriation<sup>1</sup> for the Workforce Innovation Fund (WIF), a larger initiative that involved 43 other grant awards to test new approaches in workforce services and required grantees to fund third-party evaluations (of varying levels of methodological rigor) of their projects. Similarly, the two PFS pilots were required to have grantee-funded evaluations but were also required, as per DOL's Solicitation of Grant Applications (SGA) (DOL SGAPY11-13) to use a rigorous experimental or quasi-experimental design. PFS projects opted for an experimental design, known as a randomized control trial (RCT), that required random selection of treatment and control groups, whose average outcomes would be compared in order to determine the "impacts" of the interventions.

DOL also sponsored a process study of the two PFS pilots, conducted by Abt Associates, to document project implementation and provide information on the PFS approach to policymakers and program administrators. This is the process study's final report. An earlier interim report documented the development of the pilots and their initial implementation. This final report updates information about the implementation of the pilots and includes observations based on the entirety of the DOL-supported portion, or first phase, of the PFS grants. Specifically, this report documents the services provided to participants through the PFS interventions, the implementation and management of the RCT evaluations, and the evaluation results, including whether target outcomes were met, for the first phase of the pilots.

Data sources for the process study include (1) interviews with grantee representatives and stakeholders conducted during phone calls with grantee representatives and site visits to each pilot, including small focus group discussions with program participants; (2) documents including the SGA, review criteria, grant application and contracts between pilot partners; (3) performance data and quarterly reports provided by the grantees to DOL; (4) documents about service providers including website materials and prior evaluation reports; and (5) evaluation and validation reports.

### Key Components of DOL's PFS Pilots

The key components of the PFS pilot in each state are described below.

**Government.** The Massachusetts Executive Office of Labor and Workforce Development and the New York State Department of Labor are the respective government sponsors for the pilots. In addition to the approximately \$12 million grant each received from DOL, both states committed additional resources (\$15 million in Massachusetts and \$11 million in New York) to continue the delivery of services and the evaluation of impacts. These additional state investments will significantly increase the scale of the two projects over an extended time frame and yield greater social impacts. In both states, the PFS projects are managed as a single initiative that operates in two phases.

**Social problem and target population.** Both pilots opted to address the problem of recidivism experienced by individuals who had been formerly incarcerated. In Massachusetts, the PFS pilot targeted young males (ages 17 to 24) who were on probation or parole or who are exiting the juvenile

---

<sup>1</sup> In Public Law 112-74.

justice system. In New York, the PFS pilot targeted high-risk adult parolees as they are being released from prison.

**Intermediary.** Under a PFS model, the sponsor (the government in this case) hires an intermediary to develop the PFS project, manage all aspects of it, coordinate the roles and responsibilities of all stakeholders, and secure the private sector and philanthropic financing. Massachusetts selected Third Sector Capital as its intermediary while New York chose Social Finance. Both of these organizations had experience in developing PFS projects, and both created management companies to provide oversight and coordinate all aspects of the respective projects once they became operational.

**Investors.** The two projects had to identify investors to provide up-front capital for operational and service delivery costs. A total investment of \$19.4 million was made by investors in Massachusetts and \$13.5 million in New York, with both pilots receiving funds from for-profit investors and philanthropic organizations. However, the pilots each used a different financing model, with Massachusetts relying primarily on a loan from a single investor (Goldman Sachs) and New York using one financial investor (Bank of America Merrill Lynch) to secure investments from numerous investors through a Private Placement Offering (PPO). In both states, the pilots benefitted from the funding provided by foundations (such as the Laura and John Arnold Foundation and the Boston Foundation), and performance-based investments by other non-profit investors (such as the Kresge Foundation and Living Cities).

**Evaluation and independent validator.** Both pilots opted to conduct an experimental evaluation involving an RCT. The evaluations were to measure the impacts of the pilots on participants' employment and recidivism; and these impacts would be used as the basis for determining if investors will recover their principal and the size of the return they would receive on their investment. Massachusetts opted to engage a third party to operationalize and oversee its evaluation plan, while the New York pilot used the capacity and expertise in-house to implement their approved RCT through its Research Office at the state Department of Corrections and Community Service (DOCCS). Both sites also engaged an independent validator to oversee the implementation of the evaluation, monitor its fidelity and validate the findings. The validator was responsible for developing and applying a plan to ensure the evaluation adheres to the design. Massachusetts contracted with Public Consulting Group (PCG) to serve as its validator, while New York selected Chesapeake Research Associates (CRA)/Mike Puma Associates.

**Service providers and services:** Both projects selected service providers that had been in operation for many years and had solid reputations in their field, as well as the capacity to manage data and participate in a randomized control trial. Roca, the provider in Massachusetts, had a four-year service model that involves two years of intensive services and an additional two years of post-program follow-up, with a focus on building strong relationships between the program staff and participants to change destructive thinking patterns, combined with transitional or subsidized employment, vocational and basic skills training, and work readiness training. The service provider in New York, the Center for Employment Opportunities (CEO), operated a four-to-six month program that involved training in life skills and work readiness, paid transitional employment, job placement assistance, and post-employment follow-up.

**Outcome measures and impact targets.** Each pilot established the outcome measures that would be used estimate program impacts, and formulas for calculating the level of effectiveness of the interventions and the estimated public budget savings. For their key measures, both pilots selected

post-program employment and reduction in recidivism, as determined by comparing the average outcomes for the treatment and control groups. For the employment impacts, the pilots compared employment and earnings, and estimated budgetary savings resulting from increased taxes, reduced public benefit receipt, and lower “costs” borne by crime victims. Recidivism impacts were determined by comparing the average number of “bed days” in jail or prison for treatments and controls. While each project used similar measures for employment and recidivism, the specifics of the measures varied across the two projects.

**Payment formulas for investors.** Both pilots also developed formulas for calculating payments to investors, based on the impacts achieved. While the specifics were different, both pilots had thresholds or tiers, which involved increasing levels of payment per capita at the different thresholds, for both the employment and recidivism impacts.

**Technical assistance.** While not required by DOL’s SGA, the Harvard Government Performance Lab (GPL), which had assisted with the development of other PFS projects, offered pro bono assistance to both states in designing and structuring the project and also helped develop the evaluation methodology.

### **Experiences of DOL’s PFS Pilots**

**Massachusetts and New York were both interested in PFS approaches before the release of the DOL SGA in June 2012. The DOL SGA led both states to accelerate the development of the PFS projects and formalize relationships.** Massachusetts and New York had both already explored the possibility of developing a project that used PFS and had dedicated resources to the conceptualization and design of PFS projects. PFS partners and stakeholders in both states reported that their respective governors played important roles in developing interest in and support for PFS. Building on the existing interest in PFS, stakeholders in both states acknowledged that the release of the SGA focused their interests by specifying design requirements, giving them a schedule for submitting an application, and requiring a plan for execution of the pilot.

**Both states planned to continue the pilots beyond the DOL grant period and use state funds that had previously been designated for PFS activities.** Each project had two phases: Phase I, with potential payments from Federal funds and Phase II, which added extra years to scale up enrollments, deliver services, and allow for potential payments from state funds (\$15 million in MA and \$11 million in NY).

**The pilots involved extensive planning and detailed contracts that prescribed roles and responsibilities, as well as a transparent management structure.** In each state, the key stakeholders reported that they came to the pilot with different backgrounds and interests that influenced how the pilots were developed. The key topics addressed in each state’s partnership contract were defining the intermediary’s and other partners’ responsibilities; documenting the entire evaluation plan including the specification of the experimental design and related methodologies and expected outcomes of the service intervention; and savings that could potentially be realized. Both pilots were able to implement and manage the PFS approach and address issues that arose using partnership structure that the pilots put into place.

**Both pilots operated largely as planned but had to address several challenges, particularly related to the RCT evaluations.** DOL’s SGA specified that the PFS pilots develop a partnership agreement between the four entities (government, intermediary, investors, and independent outcome



validator) to design, finance, implement, administer, monitor, and evaluate the PFS initiative. This required the partners, led by the pilot's intermediary, to develop a detailed agreement that later served as the base for a contract specifying the partnership roles and responsibilities, including the management structure, the intervention or service model, the definition of performance measures, and the evaluation design used to assess the impacts of the interventions. The pilots engaged all of the partners required by the DOL SGA, and several other partners played key roles in each pilot. The SGA required the involvement of a government agency (the grantee), an intermediary, investors, and a validator. Both pilots included providers as major partners, included specific evaluation partners in addition to validators, and benefited from considerable support from Harvard's GPL. As discussed below, most of the implementation issues stemmed from operationalizing the RCT evaluation design.

**Neither service provider made major changes to their service models, but they did seek to change aspects of their programs.** The NY pilot focused on enrolling a high-risk population shortly after release from incarceration, while the MA project sought to expand to other locations and increase the number of individuals served overall. The PFS approach requires a balance between using a program that is considered promising enough for investors to feel confident that it is likely to achieve outcomes targets but include innovations or changes that are enough of a departure from "business as usual" that the government agency would be unlikely to take on such risk. In both states, partners elected to utilize a provider long-standing service model and program but to leverage the PFS funding to increase their reach—in Massachusetts by expanding Roca's service area, and in New York by recruiting more disadvantaged individuals, particularly recent parolees assessed as being at high risk of recidivism, a group they thought could benefit from the CEO program.

**Both service providers changed recruitment procedures under PFS to meet the requirements of the RCT design, while at the same time attempting to increase the number of individuals referred to and enrolled in the programs.** Prior to PFS, Roca recruited participants by going to "where they are." This could involve recruiting youth directly in the community (e.g., circling the neighborhoods in the Roca van or on foot, knocking on doors) or within the criminal justice system as an individual prepares for release (e.g., jail, court), and through some referrals by partners. Under PFS, Roca received the names and contact information on treatment group members and then had to seek out individuals based on the contact information provided. In the New York project, parole officers had previously used their own discretion, based on their sense of which parolees would be able to succeed in the CEO program, in suggesting that parolees consider the program, and individuals made their own voluntary decision to seek it out. Under the PFS project, parole officers referred all treatment group members, who had all been assessed as being of high risk and indicated that enrollment in CEO was a "special condition" of their parole.

**The service expansion and changes to recruitment did change some aspects of how the programs were delivered.** Roca's expansion to a Boston site involved a number of challenges due to the high gang presence in the service area and among the referred participants. Roca took a number of steps to reduce incidents of conflict between members of rival gangs; this involved limiting some in-person activities and work crew assignments. Also, some participants experienced substance use disorders, which led Roca to seek out services to address those challenges. At CEO, the referral process under PFS resulted in some treatment group members viewing the program as a requirement of their probation, though many did engage with the program soon after being released from incarceration.

**Conducting the RCTs and the evaluations required organizational expertise and experience.**

Obtaining high quality and complete data from a range of sources also required time, resources, and expertise. Due to the first external evaluator not having the capacity needed to carry out all of the evaluation functions, Massachusetts ended up contracting with a larger evaluation organization. New York conducted the evaluation through an internal research office, and although the evaluation was carried out successfully, it took a great deal of staff time, and the office was not able to provide timely feedback as some partners requested.

**Neither pilot found positive impacts on the pre-specified target outcomes, and neither pilot requested a payment from DOL.** Both pilots measured impacts on recidivism and employment. Both pilots are continuing their PFS pilots and evaluations under the state-funded Phase II. Several partners reported they viewed the DOL-funded Phase I results as “interim.”

**Both the validators certified the evaluation’s findings but took different approaches.**

In Massachusetts, Public Consulting Group reviewed the evaluation findings memo and the data files they used to produce the estimates. The validator in New York, Mike Puma Associates, took on a broader role than the validator in Massachusetts and was in regular contact with pilot partners over the course of phase 1, first in establishing the validation methodology and determination process and then monitoring implementation of the project and the interpretation of results, but did not receive access to the data or report on the analysis files.

**Key Findings and Suggestions for Future PFS Pilots****Partners pointed to many positive aspects of the pilots and are optimistic about achieving outcome targets in the second phase in spite of not achieving target outcomes in the first phase.**

Launching the PFS pilots was complex and required coordination among many partners bringing different interests and experiences to the projects. Both pilots successfully brought on required partners, implemented the service model and conducted rigorous evaluations. Partners reported that in spite of the results from the first phase, the experience of designing and implementing the PFS pilot was a worthwhile learning experience.

**It was challenging to conduct the projects within the four-year timeframe of the DOL grants.**

During the four years of the grant, the projects needed to establish agreements with partners, launch the service intervention, serve participants, measure outcomes, and determine impacts. Both projects had built in a second phase beyond the period of the DOL grant, in order to serve a larger number of participants and detect larger impacts. They also noted that with a longer timeframe it may have been possible to do a pilot phase for which outcomes would not be included in the formal evaluation. For government agencies with strict funding cycles, it may be important to consider what types of interventions and outcomes are appropriate in order to be able to detect impacts within the given timeframe.

**The RCT evaluations were successfully implemented and approved by their respective validators, demonstrating the feasibility of their use in the PFS context. However, conducting the RCTs requires significant organizational expertise and experience.** RCTs are considered the most rigorous research design, and both DOL pilots opted to use them in order to ensure confidence in stated results. DOL also specified the use of an independent validator, taking another step to ensure confidence in study results. The evaluations were successfully implemented and approved by their respective validators, demonstrating the feasibility of their use in the PFS context. The use of RCTs

was not without challenges, however, as they are complex to implement appropriately and require expertise and experience in putting this type of design in place in a program environment. Based on the experience of these projects, having the government sponsor identify the evaluation standards for evaluation design and analysis to be used for the evaluations of the PFS projects (such as, for example, those for the Department of Education’s What Works Clearinghouse or DOL’s Clearinghouse for Labor Evaluation and Research). It is important to state these standards clearly at the outset of the pilots as part of the funding announcement to ensure that all parties are aware of the needs and requirements of the evaluation. Independent review of the planned evaluation design could ensure that standards are being followed and will be met in the final analysis of the results.

**If RCT methodology is a required approach, the government sponsor should be specific about design requirements and specify how results should be presented.** Based on the experiences of the two DOL pilots, a PFS sponsor should provide detailed requirements related to how final results should be presented. For future PFS projects, it is worth carefully considering ways to be transparent about evaluation plans and progress through the life of the project.

**The PFS partnership structure is a crucial aspect of a viable PFS pilot.** A number of those interviewed for this study noted that a PFS partnership structure should ensure that all key players that have a role in making the intervention function should be included in the PFS management structure. Partners observed that having a highly engaged partnership was essential for ensuring a viable project and aided in aligning interests of the various partners. Keeping the partners engaged for the entirety of a PFS pilot is challenging given shifting priorities and focus especially as state administrations change. For example, the Massachusetts state partner observed that the pilot necessitated coordination and collaboration among the partners, encouraging alignment of interests and development of innovative approaches to address the needs of the target population, calling this a “powerful feature” of the PFS approach.

**PFS projects require significant time and resource investment by the key partners.** Both pilots were able to successfully develop and manage the projects, but partners reiterated that doing so took a great deal of time. Staff changes were especially challenging over the course of the grants, as there was a steep learning curve for any new staff to understanding the project’s technically complex concepts and design issues. The projects benefitted from pro bono support and the strong commitment at the state level to exploring the PFS approach.

**Partners had different perspectives on the costs of the PFS pilots.** Observations from partners highlight that costs are likely to be a consideration for stakeholders in future PFS projects (though a cost study was beyond the scope of this process study). Both states reported that they invested significant staff time in developing, managing, and overseeing the project. Moreover, additional evaluation oversight, as discussed above, would likely increase these costs, although these costs may decrease as the government becomes more familiar with the PFS approach. Some partners reported that the time required for investors, especially philanthropic investors, to oversee the projects, represented an additional cost that may be difficult for some to absorb. In contrast, for service providers, those interviewed reported that the costs were not significantly different from the regular course of business, and that participating in a PFS pilot did not affect their program costs.

**Concluding Thoughts**

The DOL PFS pilots offer a rich set of experiences in implementing a PFS model. These experiences indicate that strong partnerships are necessary to launch and sustain such projects, and the two pilots developed and operationalized the complex partnerships needed to carry out their initiatives. However, implementing a high quality, rigorous evaluation of the interventions proved to be the most difficult aspect of the two projects, and may have contributed to the inability to detect positive impacts and pay investors under the DOL pilots. This indicates the importance of carefully planning and supporting this aspect of the PFS model in future efforts.

### 1. An Introduction to Pay for Success and DOL's PFS Grants

In September 2013, the U.S. Department of Labor (DOL) awarded two four-year grants to operate pilot projects to test the use of a Pay for Success (PFS) approach. Under PFS, financing for the initial costs of services comes from investors who are repaid, possibly with a substantial return, only when specific, pre-determined results are achieved and verified.

The PFS model was envisioned as an approach for potentially helping payors – often state and local governments – fund services and produce long-term benefits, such as reduced programmatic costs, more efficient spending, and better outcomes for people and communities. PFS potentially offers to governments and other payors a way to finance services that might not otherwise be feasible due to limited budgets. Beyond this primary goal, PFS initiatives are thought to have the potential to influence how payors will operate services in the future, by possibly creating incentives for: innovation, effective preventive or proactive services, flexible service delivery, and adaptable management structures.

The DOL pilots were the first effort by a federal agency to support the PFS model. However, the first project testing Social Impact Bonds—a model with similar principles as PFS—was launched in September 2010 at the Peterborough Prison in United Kingdom. Since that time, many PFS projects have been launched in the US and in other countries, and the PFS field has evolved considerably to cover a range of policy areas, including early childhood, employment, criminal justice, and homelessness.

The DOL grants were awarded to two state workforce agencies: the Massachusetts Executive Office of Labor and Workforce Development (EOLWD) and the New York State Department of Labor (NYSDOL) and were for approximately \$12 million each. Both pilots focus on improving employment outcomes and reducing recidivism among individuals newly released from incarceration. The Massachusetts pilot targeted young male parolees and probationers, while the New York pilot focused on high-risk adult parolees.

Funds for the PFS pilots were drawn from an appropriation<sup>2</sup> for the Workforce Innovation Fund (WIF), a larger initiative that involved 43 other grant awards to test new approaches in workforce services and required grantees to fund third-party evaluations (of varying levels of methodological rigor) of their projects. Similarly, the two PFS pilots were required to have grantee-funded evaluations but were also required, as per DOL's Solicitation of Grant Applications (SGA) (DOL SGAPY11-13) to use a rigorous experimental or quasi-experimental design. PFS projects opted for an experimental design predicated on a well-executed randomized control trial (RCT). This design required random selection of treatment and control groups, whose average outcomes would be compared in order to determine the "impacts"<sup>3</sup> of the interventions.

---

<sup>2</sup> In Public Law 112-74.

<sup>3</sup> This report uses the term "impacts" to identify the positive or negative results calculated by comparing average results for treatment and control groups, as per common terminology in most social science research. This differs from the term used in the Solicitation of Grant Applications for the PFS projects, in which these results were identified as "outcomes."

Both states planned to continue the pilots beyond the DOL grant period and use state funds that had previously been designated for PFS activities. Each project thus had two phases: Phase I, with potential payments from Federal funds and Phase II, which added extra years to scale up enrollments, deliver services, and allow for potential payments from state funds (\$15 million in MA and \$11 million in NY).

### 1.1 The Pay for Success Approach

Since the introduction of PFS/SIB model in 2010, there has been a great deal of literature generated about the approach, and several organizations have centers and/or staff dedicated to understanding and supporting PFS projects. Descriptions of PFS highlight the innovative aspects of the approach, such as the model's ability to pay only for services that lead to positive outcomes, shift risk from the government to private investors, promote evidence-based policy, and potentially scale-up services that are proven to be effective.

A primary goal of any PFS effort is to promote effective solutions to social or environmental problems. While each pilot is unique in the target population, financing vehicles, partners, and the intervention or services provided, the PFS financing and managerial structure is typically designed around the stakeholder roles and responsibilities described below.

**Government.** The payor, often a federal, state, or local government agency, is responsible for selecting a social problem to address through the PFS model. The payor also is responsible for (1) contracting with an intermediary or transaction coordinator to manage key aspects of the project, (2) managing the funding for repaying investors, and (3) helping determine the outcome measures, targets, and potential savings to the government.

**Intermediary.** This third-party organization, contracted by the government or other payor, manages all aspects of the PFS initiative, coordinating the roles and responsibilities of all stakeholders and securing financing through private or philanthropic investors.

**Investors.** Not-for-profit or for-profit entities, or individuals, who provide financing for the PFS effort with the expectation that the government will repay the investment based on the achievement of predetermined outcomes. Typically, this financing covers the costs associated with the delivery of the service intervention.

**Independent Validator.** The party or parties responsible for measuring the impacts of the service intervention. These metrics become the basis for determining if and how much investors are repaid. The payor or government often contracts with an external third-party to verify the fidelity of the evaluation design and results.

**Service provider.** This organization provides the intervention designed to achieve a specific set of social and/or economic outcomes.

**Target population.** The group of individuals requiring the identified services and that are recruited and enrolled to participate in the services.

### 1.2 The Approach to PFS Used by DOL

DOL's PFS SGA emphasized that in addition to funding services that would provide improved workforce outcomes, their interest in the model also centered on testing the approach and understanding its feasibility. (DOL PFS SGA 2011, 1) The SGA included the following statement:

*In piloting the Pay for Success model, the Department will provide funding for one or more projects that will demonstrate the feasibility and viability of this funding model for providing positive workforce outcomes. The Department's objectives in establishing this pilot are to:*

*Test a model for government investment in preventative and innovative service delivery models that transfers risk to the private sector.*

*Learn whether the Pay for Success concept is feasible in the workforce development policy arena.*

*Determine whether preventative social services complementing workforce development programming "pays off."*

The SGA required that prospective grantees have a fully formed partnership in place, consisting of a state, local or tribal government agency, intermediary, investors, and independent outcome validator, and that the partners have a partnership agreement in place. Consistent with the PFS field at the time, the DOL SGA identified the key partners and their responsibilities, as follows:

- **The government agency** which was the grant applicant and the entity responsible for a) identifying and entering into a contractual agreement with the intermediary in which payment would be made for the achievement of specific target outcomes, and b) submitting necessary reports (such as quarterly reports and information on impacts and payment) according to the provisions of the DOL grant agreement.
- **The intermediary** who was required to enter into the contract with the state agency, and who was responsible for coordinating the Pay for Success strategy; achieving the negotiated outcome(s) for the target population by contracting with service delivery providers; raising funds from investors to cover the operating costs of achieving the outcomes; and collecting and sharing data with the government grantee to fulfill the grant agreement. The intermediary also had flexibility to change or modify the service provider's delivery methods and even to change providers, as noted above.
- **The investor(s)** who were defined as not-for-profit or for-profit entities that sought to invest in promising social service interventions and who "accepted the inherent risk of total non-payment (writing off the operating funds as a loss) if the target outcome(s) are not achieved, but are willing to invest based on their confidence in the proposed intervention, and a guarantee from the grantee that they will be paid for outcomes (i.e. earn their principal investment back plus a return on investment) should the target outcome(s) be met." Investors were to agree to fund all of the operational and service delivery costs of the intermediary and service provider(s) throughout the lifetime of the proposed project.
- **The independent validator** who was to be under a contract with the government grantee which was to pay the validator throughout the pilot. The validator was responsible for informing (and agreeing) to the evaluation methodologies, checking the data files and analyses, determining

---

## AN INTRODUCTION TO PAY FOR SUCCESS AND DOL'S PFS GRANTS

---

whether the outcome (i.e., impact) targets were met, and providing documentation that would be used in reporting to DOL and in determining payments due to the grantee.

DOL specified that service providers should be involved in the pilot but did not name them as a required partner, reflecting that the intermediary and other partners would be responsible for closely managing the services and making adjustments to service provision as necessary.

Also, while the SGA required that there be independent validator responsible for verifying outcomes, it did not require an evaluator who would conduct and coordinate the randomization process; collect and analyze data, including use of statistical methods; and to determine estimated impacts, error rates, and statistical significance. Nonetheless, both pilots opted to engage a separate organization to perform those evaluation tasks.

Finally, it should be noted that the technical proposals in the grant applications became part of a grant agreement between each state and ETA. The grant agreement had administrative and financial provisions, including a requirement for grantees to provide quarterly reports documenting grant progress and data on enrollments and performance. ETA also provided programmatic technical assistance to grantees and partners under the WIF technical assistance contract with Maher and Maher/Jobs for the Future. Evaluation technical assistance on data and statistical issues was provided to ETA by Abt Associates, as part of its role as the National Evaluation Coordinator for WIF grants.

Based on the SGA, DOL anticipated that launching and implementing a PFS project would be a complex endeavor, and that the pilots would be of interest to an emerging field and to state, local and tribal governments interested in the approach. In particular, at the time that DOL released the SGA and awarded the grants, it was not clear how feasible it would be to bring together the required partners and develop the contractual agreements needed to launch and sustain a PFS project. As such, the DOL PFS grants were intended to not only provide services to program participants, and also help the field understand more about putting PFS into practice.

### 1.3 The Pay for Success Field since the Launch of the DOL Pilots

Since the DOL grants were awarded and launched in 2013, the PFS field has evolved considerably. By the end of 2015, five additional PFS projects were launched that focused on a range of policy areas, including early childhood, employment, criminal justice and homelessness; by 2017 this number had increased to 15 and included projects focused on K-12 education, employment, criminal justice, child welfare, and mental health. (Golden et al. 2015, 1)

Several projects have released some results in terms of whether the projects have met their targeted outcomes and made payments to investors, and these results have been mixed. These projects have yielded lessons for program administrators, researchers, policymakers and other interested stakeholders that have shaped thinking on PFS.

When payments to investors have not been made because the specified outcomes were not achieved, projects may still be viewed by some observers as demonstrating successful implementation of the PFS model: the government was able to take a risk on providing promising services without bearing the cost of services that ultimately were not found to achieve pre-determined results. At the same time, projects that have met goals for improved outcomes and have had payments made have been subject to some scrutiny due to a lack of confidence in evaluation findings and the significant



repayment amount to investors. Below we describe the projects that have released some results (both interim and final), the lessons and implications based their experiences; and how the field has evolved in light of these findings since the DOL pilots were launched.

### 1.3.1 Results from Other PFS Projects

The section provides a brief summary of key PFS projects that have reported results to date.

**The HM Peterborough Prison project**, launched in 2010 in the United Kingdom, was the world's first Social Impact Bond demonstration. The project aimed to reduce recidivism among men who served short sentences (less than 12 months) at HM Peterborough Prison through an intervention that begins prior to and continues after release. The project intended to include three cohorts of 1,000 prisoners each and to operate over seven years. The target outcome was to reduce the rate of reconviction across cohorts by 7.5 percent with the possibility of early payments if the project reduced the rate of reconviction for a single cohort by 10 percent. To determine the program's effect, the evaluation used a comparison group of prisoners released from short sentences within the same time period, a quasi-experimental methodology called Propensity Score Matching. (Disley et al. 2014, 11-12 and Social Finance July 2017, 4)

In the end, the project served two cohorts, and while neither met the 10 percent recidivism reduction, collectively they did meet the 7.5 percent threshold. As a result, the investors received a single payment of their initial capital plus an amount that represented a return of just over 3 percent per year for the period of investment. The project was terminated after two cohorts because the government rolled out national reforms to probation which introduced mandatory supervision for all short-sentenced offenders. This resulted in concern about a duplication of services and the program continued to operate, but under a fee-for-service funding arrangement with results for the third cohort not included in the payment determination. A final process evaluation (Dilsey et al, 2015) conducted for the UK government report noted that there were a number of innovations in the pilot, including flexibility in the funding and resultant adaptations in the services in response to local conditions and participant needs, but that most of the innovations identified were not necessarily a result of SIB financing as other (non-SIB funded) initiatives had exhibited similar characteristics.

**The New York City ABLE Project for Incarcerated Youth** launched in February 2013. The project's goal was to reduce recidivism among 3,400 minority youth who stayed at Rikers Island jail for more than four days through an intervention that increases moral reasoning. The target outcome was to reduce the number of days youth spend in jail after being released from jail. The project used a quasi-experimental evaluation design that compared the bed days for the treatment cohort to a cohort entering custody prior to the start of the intervention. (Golden et al. 2015, 12-13)

The early results in August 2015 (reported in year three of the four-year evaluation) found that the project failed to reduce recidivism by more than nine percent (the target reduction) after one year of follow-up. As a result, the partners ended the program early and did not request repayment. In a study of the project, the intermediary described two key challenges that likely affected why the project did not meet its proposed targets. First, only nine percent of the cohort members completed all stages of the intervention, which was lower than expected. This was partly due to the security-related events occurring at the prison, which disrupted the program's courses, and also because the length of stay of participants awaiting trial and the requirements for youth post-release are unpredictable and thus interrupted participant progress in the program. Additionally, the intermediary staff reported the

amount of stress experienced by the inmates limited their ability to benefit from the program. (Berlin 2016, 6)

**The Utah High Quality Pre-Kindergarten project** was launched in 2013 with outcomes being measured for four cohorts of preschoolers over 12 years. The project is testing a curriculum that intends to increase school readiness and academic performance among low-income pre-school students, thus reducing the proportion requiring special education or remedial education services. The project is using a nonexperimental evaluation, and outcomes are based on the number of children requiring special education or remedial services each year between kindergarten and sixth grade. (Blum et al. 2016, 13)

Preliminary results were released in 2015. Initial payments to investors were made when 109 out of 110 students (99 percent) did not enroll in special education kindergarten (and the evaluation did not include a comparison group). However, this level of success raised concerns about the validity of the outcomes. The evaluation's study group was defined as those who scored less than a specified score on an aptitude test—these students were determined to be likely to be placed in special education. Early-education researchers questioned this eligibility criteria because the test used was not typically used on its own to screen for special education needs. Specifically, because a high proportion of the children enrolled in the program did not come from English-speaking homes, the low test scores may have been indicative of low English proficiency, not a learning disability, and thus the students may not have been in need of special education. Some researchers argued that the project overstated the program's effects due to this potential inaccurate assumption about the need for special education among the study group. In addition, the project received further criticism because of the large size of the payment to investors relative to the cost of the program.<sup>4</sup>

In response, the program issued a brief to provide additional information on how the children for the study were identified, the use of the aptitude test, particularly for English learners, and the reasons for not conducting a more rigorous evaluation, particularly a randomized controlled trial (RCT), that discussed cost, the difficulty, and ethics of securing a control group, and the consistency of past evaluation results. (Innocenti, no date, 2 -6)

**The Chicago Child-Parent Center PFS project** began in October 2014 and consists of four cohorts of preschoolers who are being evaluated over 17 years. The intervention provides supports to students at their preschools as well as parent support services. The project's target outcomes are to reduce the number of children who need special education after preschool (measured annually through 6<sup>th</sup> grade), increase kindergarten readiness (measured once at the end of preschool), and increase third grade literacy rate (measured once after third grade). The quasi-experimental design involves a comparison group of children who did not attend preschool.

The early results of the project were released in May 2016, with an increase in the proportion of first cohort deemed ready for kindergarten that met the established target goal. This triggered \$500,000 in payments to investors. (Eldridge et al. 2016, 1) In 2017, evaluators released findings for the second

---

<sup>4</sup> Investors received their money back with a high rate of return, at five percent interest, as long as half of the children were not enrolled in special education. An analysis in the New York Times showed that this was more than the state would have paid if it had borrowed the money on the bond market. (Propper. 2015)

cohort, which resulted in a second payment of approximately \$900,000 based on the same measure. (IFF 2017, 1 – 3) While the project has shown early success, observers have noted the weakness of the quasi-experimental methodology compared to a RCT, the perception that the intervention had already been found to be successful in other studies and did not represent an innovation or a substantial risk to investors, and the concern that the program receiving referrals for children who did not have a sufficient need for special education. (Eldridge et al. 2016, 1; Sanchez 2016, 1)

### 1.3.2 Current State of the PFS Field

In addition to results from PFS projects discussed above, the PFS field has generated a considerable amount of additional literature focused on project development and best practices. Early literature on PFS largely focused on the concept, potential and financial risks of PFS. After 2014, attention turned to lessons learned from the early PFS projects and the application of PFS to a wider array of policy areas.<sup>5</sup>

The growth of PFS has also spurred a concurrent criticism of the approach.<sup>6</sup> These concerns include that: (1) PFS projects may be more costly for the government due to the costs of legal services, evaluation, program administration and loan management; (2) the interventions considered may be too narrowly focused on those with outcomes that can be easily measured, given the difficulty in measuring outcomes in a way in which all partners can be confident in the results; (3) the interventions may not be sufficiently innovative because investors will choose to limit risk by only choosing programs with a strong evidence base; and (4) the government may still retain risk because it could be difficult to discontinue the program even if the PFS financing is discontinued.

The study based on the Riker's Island PFS project reported on difficult issues that must be resolved in PFS projects generally. These include tensions around the appropriate level of risk for such projects, given investors' tolerance for risk and the returns the government is willing to offer; the need for clearly specified outcomes upfront (and a commitment to serve given number of participants in a fixed time span) that make adaptation to unforeseen developments difficult; the appropriate level of evidence to be required for outcome payments; and the rigor required in PFS evaluations. (Berlin 2016, 7-12) However, this study also observes that "while the intervention may not have worked, the [PFS] structure can still succeed: the government does not pay for ineffective services, both the nonprofit service delivery organization and the government get the flexibility to innovate and learning what does not work can help point the way to what might." (Berlin 2016, 3). The government and

---

<sup>5</sup> For a summary see: State of PFS Field II: Emerging Literature, Updates and Tools, April 2016, p. 4-5.

<sup>6</sup> For example, a report released by In the Public Interest and the American Federation of State, County, and Municipal Employees (along with other advocacy groups) that points to a number of concerns. American Federation of State, County and Municipal Employees, Center for Effective Government, Economic Opportunity Institute, In the Public Interest, Keystone Research Group, Minnesota Council of Nonprofits, Oregon Center for Public Policy, & Refund America Project. (2015). A Guide to Evaluating Pay for Success Programs and Social Impact Bonds. Washington, D.C.: In the Public Interest. Accessed at: <https://www.inthepublicinterest.org/wp-content/uploads/ITPI-Pay-for-Success-Guide-Dec-2015.pdf>

project partners were able to experiment with the approach, and the government did not end up paying for services that did not lead to desired outcomes.

A common issue among the completed projects, especially those that were found to achieve outcomes, was how to select the appropriate degree of rigor for the evaluations used to determine whether outcomes were achieved and payments should be made. The evaluation design and approach for PFS projects continues to be a matter of discussion and debate within the PFS field. RCTs, the “gold standard” in evaluation because of the ability to provide a causal link between programs and impacts can be expensive and difficult to implement and are still subject to scrutiny based on their specific research designs. Furthermore, past studies have shown the replication of positive results can be difficult, raising questions about the ability of new projects to produce positive impacts in different settings and environments.<sup>7</sup>

PFS projects have evolved over time, with varying degrees of rigor. Some have used more rigorous evaluation methods and tighter administrative elements, as evident in the 2018 legislation authorizing the Social Impact Partnerships to Pay for Results Act (SIPPRA) Demonstration Projects under the U.S. Treasury Department, whose *Notice of Funding Availability*,<sup>8</sup> published in 2019, included new requirements for demonstrating savings and additional oversight.

### 1.4 Process Study Research Questions and Data Sources

To document the design and implementation of the two PFS pilots, DOL sponsored a process study, conducted by Abt Associates. A previous interim report focused on describing the early implementation and operation of the PFS pilots, and initial observations based on data collected as the pilots were getting under way (Abt 2016). This report updates information about the implementation and payment outcomes for the DOL-supported portion, or first phase, of the PFS pilots.

The interim report addressed the following research questions:

- **How were the PFS pilots designed, and what were the national and local contexts in which they were developed?** The study examines the political, policy, and program environments in which grantees and their partners developed their pilots.
- **How did grantees identify, recruit, and maintain partnership organizations for their PFS pilots?** The study documents changes made in the partnerships and responsibilities over time as well as specific challenges and issues faced by individual partners.
- **How is the PFS initiative managed and operated?** The study examines the role of the state grantee and the intermediary organization in managing the project, facilitating partner communication, handling dispute resolution, addressing investor issues, and overseeing service provision.

---

<sup>7</sup> See for example Miller, C.; J. Bos; K. Porter; F.M. Tseng and Y. Abe (2005). *The Challenge of Repeating Success in a Changing World: Final Report on the Center for Employment Training Replication Sites*. MDRC: New York.

<sup>8</sup> Found at: <https://home.treasury.gov/system/files/226/SIPPRA-NOFA-FINAL-FY2019.pdf>.

- **How was the PFS financing structure developed and operated?** The study examines the identification and integration of private and philanthropic capital for the pilots; the financing mechanisms used to pay for the services; and outcome measures, targets and payoff mechanisms used to provide returns to investors.

This report updates findings for the questions listed above, and also focuses on addressing the following two research questions that are informed by understanding the progress of the pilots over the entirety of their first phase of operations.

- **How does the service delivery operate and what services were provided to participants?** The study documents the nature and content of services provided, including participant recruitment and tracking of participant outcomes.
- **Were the target outcomes achieved and payment to investors made?**
- **What are the challenges and promising practices experienced by the PFS projects?**

Specifically, this report documents the experience of the service providers over time, the implementation of the randomized control trial evaluations and the validation process, and the final outcomes for Phase I of the pilots.

The data sources for the process study were: (1) interviews with grantee representatives and stakeholders conducted during phone calls with grantee representatives and site visits to each pilot, including small focus group discussions with program participants; (2) documents including the SGA, review criteria, grant application and contracts between pilot partners; (3) performance data and quarterly reports provided by the grantees to DOL; (4) documents about service providers including website materials and prior evaluation reports; and (5) evaluation and validation reports.

This report draws largely on information collected from site visits conducted in the summer and fall of 2017, supplemented with information from the interim report that drew from site visits conducted in fall 2014. During both rounds of site visits the research team used interview protocols that were organized around topics and questions developed in conjunction with DOL. The key topics reflect the research questions discussed above and include the establishment and maintenance of the partnerships, recruitment of investors and development of the financing structure, design of the RCT, the outcome measures, targets and structure of the payments, and the nature of the service intervention. Semi-structured interviews were conducted with representatives of the intermediary, the service provider, investors, key state agencies, executive-level leadership, the evaluator, the validator, and technical assistance providers for each pilot. Unless otherwise noted, observations and opinions about these PFS projects are taken directly from data collected through the structured interviews. (See Appendix A for a list of organizations interviewed).

### 1.5 Structure of the Report

The remaining chapters of this report describe the implementation and execution of the two DOL PFS pilot projects, as well as lessons we can draw from their experience. Chapter 2 describes the design and operation of the pilots, including the roles of partners and the financing strategy. Chapter 3 describes the service providers and the service interventions for the two pilots. Chapter 4 discusses the challenges in conducting random assignment and achieving target enrollments by the two pilots,

## AN INTRODUCTION TO PAY FOR SUCCESS AND DOL'S PFS GRANTS

---

and Chapter 5 describes the evaluation results and validation of outcomes by the external validator. Finally, Chapter 6 summarizes key findings that provide suggestions for future PFS projects.

## 2. The Design and Implementation of the PFS Pilots

This chapter first describes how each state designed and implemented the PFS pilots, including discussion of the formation of partnerships, contracts, the management structure amongst partners, and the financing arrangements. The remainder of this chapter provides an overview of the evaluation designs for each pilot and how outcome targets were defined.

### 2.1 Launch and Early Operation of the PFS Pilots

As noted in Chapter 1, part of DOL's intent in awarding the grants was in understanding the feasibility of PFS projects. These two pilots were among the first PFS projects launched in the United States, and it was not clear at the outset how straightforward it would be to bring all of the required partners on board to implement the pilots. To address this issue, this section describes the early development and planning for each of the pilots.

#### 2.1.1 PFS Model Development Prior to SGA Release

In both Massachusetts and New York, PFS efforts were underway before the DOL SGA announcement, and as described below, the application deadline in the SGA accelerated the development of the projects and formalized those relationships in both states. Specifically, before DOL released its SGA in June 2012, both states had already explored the possibility of developing a project that used PFS and had dedicated resources to the conceptualization and design of PFS projects. PFS partners and stakeholders in both states reported that their respective governors played important roles in developing interest in and support for PFS.

The overall interest of the public sector in evidence-based policy generally, and PFS specifically, also contributed to the development of the pilots. Because of the interest, the Harvard SIB Lab (now known as the Government Performance Lab or GPL) was launched in 2011 and provided of support to conceptualize and implement the two pilots. Details of these early activities for each state are described below.

#### Massachusetts

PFS partners in Massachusetts reported that the initial discussions regarding whether and how a PFS approach could work in the state began in 2010 between then-Governor Deval Patrick and members

#### Designing the PFS Pilots: Key Points

- *Massachusetts and New York were both interested in PFS approaches before the release of the DOL SGA.*
- *The DOL SGA led both states to accelerate the development of the PFS projects and formalize relationships.*
- *Executive-level commitment in the states helped the pilots move forward; the projects also enjoyed the support of advocates for the PFS approach, especially the Harvard Government Performance Lab (GPL).*
- *The pilots engaged all of the partners required by the DOL SGA, but also included service providers and evaluators, in addition to validators, and benefited from support from Harvard's GPL.*
- *The pilots required extensive planning and detailed contracts that prescribed roles and responsibilities, as well as a transparent management structure.*
- *The states took different approaches to financing and had different funding structures.*
- *Both states opted for RCT designs, which required decisions around sample size, data sources, statistical methods, and back-up methodologies.*

of his cabinet, including the secretary of Administration and Finance and the commissioner of Youth Services. The concurrent formation of the Harvard GPL at Harvard's Kennedy School of Government also generated interest in the approach. As a result of the activities of these stakeholders, in 2012 the Massachusetts legislature created the Social Innovation Financing Trust Fund. This fund authorized and provided \$50 million to the state's secretary of Administration and Finance and state agencies to enter into multi-year PFS contracts. State staff interviewed for this study reported that this commitment to uninterrupted, longer-term funding was important to attracting private capital since payments would not be subject to annual appropriations risk.

After creation of the fund, state staff moved ahead with planning a PFS project. The Office of Administration and Finance issued a Request for Response (RFR) to identify potential policy focus areas and find an intermediary that could both guide the planning process and engage appropriate partners. Based on responses to the RFR, the state contracted with Third Sector, an organization specializing in capacity building for nonprofits. Third Sector's response to the RFR was submitted in collaboration with several potential partners that were well positioned to engage investors and deliver services, depending on the programmatic direction selected by the state. By the time DOL released its PFS SGA in June 2012, Massachusetts was already identifying potential PFS projects and searching for an intermediary to pursue the grant opportunity.

### **New York**

Like Massachusetts, New York had an interest in the PFS model before the announcement of the DOL PFS grant opportunity in 2012. Initial interest in PFS was connected to a broader policy commitment to public safety. According to state agency staff, Governor Andrew Cuomo and the state's executive leadership were committed to improving quality of life through a reduction in crime. One of Governor Cuomo's public safety initiatives—the Work for Success initiative—sought to reduce recidivism by providing job-readiness services and occupational skills training for individuals transitioning from incarceration. During this period, New York also was focused on developing evidence-based investments. The Work for Success initiative, for example, was directed to “evaluate the employment programs that currently exist for the formerly incarcerated and carry out evidence-based, action-oriented research to identify which strategies work best.” (New York State Press Release, 2012, 1)

New York State also had an interest in performance-based financing that pre-dated the DOL SGA. In part this resulted from New York City's Social Impact Bond Project at Rikers Island Jail that was implemented prior to the DOL SGA (see Chapter 1). As a result of the proximity and prominence of this effort, coupled with Governor Cuomo's interest in connections between recidivism and public safety, in 2013 the state legislature provided up to \$30 million in funding for PFS initiatives in a range of areas, including health care, child welfare, early childhood development, and public safety.

In retrospect, the New York PFS partners interviewed for this study reported that the DOL SGA accelerated existing momentum by providing resources to launch a PFS pilot. The partners also indicated that New York's pursuit of a PFS model would have continued without the DOL grant, although at a slower pace.



### 2.1.2 Release of the DOL SGA, Two-Phase Implementation, and Selection of Partners

The DOL's SGA for the PFS pilots was released in June 2012, and applications were initially due in early December of that year. As discussed above, stakeholders in both states acknowledged that the release of the SGA focused their interests by specifying design requirements, giving them a schedule for submitting an application, and requiring a plan for execution of the pilot.

DOL's SGA largely reflected the perspective of the field about the PFS approach and how such projects should be structured. Specifically, DOL required that the following PFS elements be in place when the application was submitted (DOL PFS SGA, 2011, Section IB):

*A well-defined problem and associated target population that faces documented employment- and training-related challenges;*

*A partnership agreement that includes an intermediary, at least one investor, and an independent validator;*

*A service intervention that addresses the needs of the target population and can achieve predetermined outcome targets, one of which must be employment;*

*A financing model that details investors' up-front financing of pilot operations, demonstrates potential government savings, and determines the outcome targets that trigger payment to investors; and*

*Plans to implement an experimental or quasi-experimental evaluation methodology to determine if outcome targets have been met and to ensure that impacts can be confidently attributable to the service intervention.*

Massachusetts and New York both focused their PFS applications on the "social problem" of recidivism among formerly incarcerated individuals and others involved with the criminal or juvenile justice systems. In addition to being priority policy areas in both states, addressing recidivism fit well with the PFS approach, according to partners from both pilots, because the potential outcome metrics would be relatively straightforward and measurable, thereby facilitating payment based on the achievement of specific targets. In addition, the applications presented hypothesized budgetary savings that were intended to be clearly identifiable and could potentially be tracked over a relatively short time period. Moreover, the partners reported that the focus on improving employment and earnings fit well with DOL's overall objectives for the pilots, and as noted earlier, both pilots opted for an experimental design.

**Two phase pilots.** Both states elected to divide their PFS projects into two phases. The first phase (Phase I) was funded with the federal grant from DOL (Massachusetts was awarded \$11.7 million and New York was awarded \$12 million) and operated from September 2014 to September 2017, though services ended under the DOL grant in late 2016. The second phase (Phase II) of the projects was funded by the states and began services in 2016 to on-board and new participants, and was slated to continue for several years more. In Massachusetts the state grant was for \$15 million; in New York it was for \$12 million.

The two-phase structure was developed for two reasons. First, a goal was to increase the number of participants in order to increase the overall potential level of government savings that could be realized from the pilots (see Chapter 5). Second, as discussed below (see Section 2.2), the RCTs that

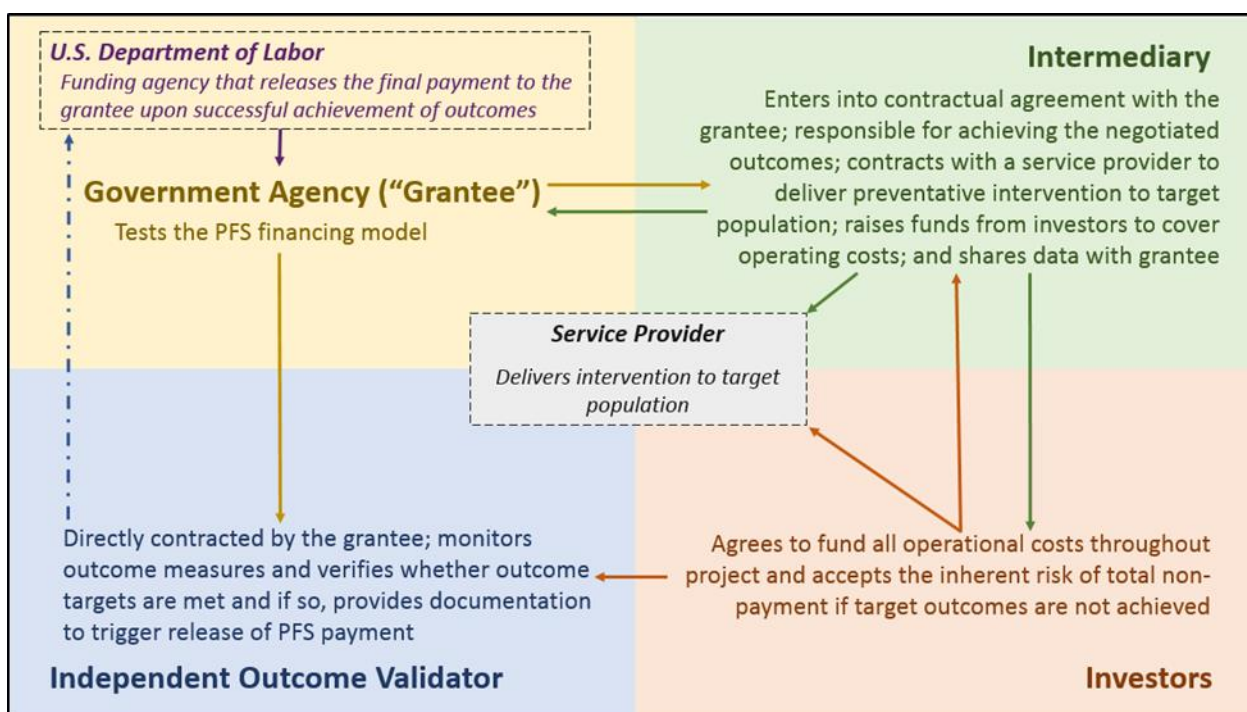
## THE DESIGN AND IMPLEMENTATION OF THE PFS PILOTS

projects chose to implement required large sample sizes to detect meaningful impacts, so increasing the number of participants strengthened the evaluation as well.

Each state’s commitment of resources to support Phase II increased partner involvement and accountability in achieving the target outcomes. Both pilots were designed so that the federal component could stand on its own; that is, the pilots were structured so partners could observe outcomes within the timeframe of the DOL grants and meet all other grant requirements. However, the partners reported that they viewed both phases as part of a single project and factored the larger sample size and longer follow-up period that Phase II allowed into their calculations of the risks and benefits of participating.

**Selection of Partners.** As outlined on Exhibit 2-1, a key requirement of the SGA was for applicants to establish a partnership that included an intermediary, at least one investor, and an independent validator. Both projects also included a service provider, evaluator, and technical assistance providers. This exhibit summarizes the roles and responsibilities of each partner and their connection to each other.

**Exhibit 2-1: PFS Pilot Partnership Structure Required by DOL’s SGA**



While the implemented pilots fulfilled the requirements of the DOL SGA and brought all of the required partners on board, there were additional partners who were key to implementation of the pilots, and there were ways in which the pilots interpreted the roles of partners differently. The partners selected for each pilot are summarized in Exhibit 2-2 below.

## THE DESIGN AND IMPLEMENTATION OF THE PFS PILOTS

**Exhibit 2-2: DOL PFS Pilot Partners**

Partners	Massachusetts	New York
<b>Government</b>	<ul style="list-style-type: none"> <li>• DOL award of \$11.7 million to the Massachusetts Executive Office of Labor and Workforce Development</li> <li>• \$15 million state investment in Phase II</li> <li>• Application submitted by Executive Office of Labor and Workforce Development; managed by Executive Office of Administration and Finance</li> </ul>	<ul style="list-style-type: none"> <li>• DOL award of \$12 million to the New York State Department of Labor</li> <li>• \$11 million state investment in Phase II</li> <li>• New York State Department of Corrections and Community Supervision designated to identify and recruit individuals for project services.</li> <li>• Pilot established an interagency MOU.</li> </ul>
<b>Intermediary</b>	<ul style="list-style-type: none"> <li>• Third Sector Capital Partners served as intermediary; created Youth Services, Inc. (YSI) to manage the contract.</li> </ul>	<ul style="list-style-type: none"> <li>• Social Finance served as the intermediary and created Social Finance-New York, to manage the contract.</li> </ul>
<b>Investors</b>	<ul style="list-style-type: none"> <li>• Goldman Sachs provided largest loan and will be paid first for any return on investment.</li> <li>• Additional loans from two nonprofit junior lenders: Kresge Foundation and Living Cities.</li> <li>• Philanthropic grant donations from the Laura and John Arnold Foundation, New Profit Inc., and the Boston Foundation</li> <li>• Roca, the service provider, deferred payment</li> </ul>	<ul style="list-style-type: none"> <li>• Over 40 high net-worth private investors and organizations through a private placement offering by Bank of America Merrill Lynch.</li> <li>• Two nonprofit lenders: Kresge Foundation and Living Cities.</li> <li>• Philanthropic donations from the Robin Hood Foundation and the Laura and John Arnold Foundation</li> <li>• Rockefeller Foundation offered a loan guarantee to reduce investors risk to 90 percent if impact targets were not met</li> <li>• Accredited investors are subject to Securities and Exchange Commission standards and are unable to publicly disclose information about the project.</li> </ul>
<b>Evaluator</b>	<ul style="list-style-type: none"> <li>• Independent evaluation contract awarded first to Sibalytics. Urban Institute replaced Sibalytics in November 2014</li> </ul>	<ul style="list-style-type: none"> <li>• The NY DOCCS Research office had responsibility for implementing random assignment and collecting the data.</li> </ul>
<b>Validator</b>	<ul style="list-style-type: none"> <li>• Public Consulting Group</li> </ul>	<ul style="list-style-type: none"> <li>• Mike Puma Associates</li> </ul>
<b>Service Provider</b>	<ul style="list-style-type: none"> <li>• Roca, which had been engaged with project partners since before DOL SGA</li> </ul>	<ul style="list-style-type: none"> <li>• Social Finance interviewed five established service providers that target the ex-offender population and selected two to participate in an invitation-only RFP. CEO was selected.</li> </ul>
<b>Technical and legal assistance providers</b>	<ul style="list-style-type: none"> <li>• Harvard Government Performance Lab helped design and structure the project and evaluation methodology and payment formulas (pro bono)</li> <li>• DOL WIF technical assistance providers, Maher and Maher and Jobs for the Future.</li> <li>• Pro bono legal assistance from Nixon Peabody LLP, Ropes &amp; Gray LLP, Goulston &amp; Storrs, P.C., and Goodwin Procter LLP.</li> </ul>	<ul style="list-style-type: none"> <li>• Harvard Government Performance Lab helped to design and structure the project and evaluation methodology (initially pro bono and then through a subcontract)</li> <li>• DOL WIF technical assistance providers, Maher and Maher and Jobs for the Future.</li> </ul>

**Service Providers.** The SGA required that service providers be involved but as noted above did not name them as a required partner. In both pilots however, the service providers were brought on board early in the project’s development as key partners. In Massachusetts, Roca deferred a portion of its payment, making the organization an investor in the project, and Roca was identified as a partner with the intermediary in the partnership agreement. Roca had a previously established relationship with the intermediary; Third Sector and Roca had jointly responded to a state-issued “Request for Information” in May of 2011 (before the release of the DOL SGA) seeking external support to identify policy areas and potential interventions for a PFS project.

In New York, the Center for Employment Opportunities (CEO) was the service provider in the initial proposal to DOL. Social Finance had selected CEO after conducting a literature review, interviewing five providers, and inviting two to submit proposals. A previous experimental evaluation of CEO was cited by project partners as one of the reasons that state leadership was interested in taking on a PFS approach in this field. Both CEO and Roca had long histories working in the criminal justice field and had the organizational capacity to implement the projects and their accompanying RCT evaluations SGA (see section 2.2 for a discussion of evaluation designs).

### 2.1.3 Development of Contracts to Manage Pilots

As described in this section, establishing agreements around roles and responsibilities, and establishing working relationships was a major part of planning for the PFS pilots, both during the grant application period and immediately after the grant award was made by DOL. In each state, the key stakeholders reported that they came to the pilot with different backgrounds and interests that influenced how the pilots were developed. Many of the partnerships required by PFS had to be formed from scratch, while others had to be revisited given the unique nature of the PFS approach. PFS stakeholders reported that they had few established templates to work from and even those who had worked successfully together in other settings were working with new concepts, relationships, time lines, and incentives. Moreover, the partners and stakeholders brought different perspectives and goals, which were reflected in the nature and substance of the PFS contracts.

Key factors that affected the development of the partnership agreement required by the SGA (and which were then developed into contractual documents) were the presence of public and private priorities and the requirement for adequate resources. State agency staff from both pilots reported in interviews conducted for this study that while they were interested in innovative ways to finance needed services under the projects, they did not want the initiative to appear to create an overly profitable investment opportunity for the private sector. The private investment partners reported that they mainly saw the opportunity as a way to tap into market demand for “impact” investing opportunities with a social purpose. The investment partners also sought a return big enough to generate interest among other potential investors. Finally, in spite of their interest, most partners were unfamiliar with working through an intermediary and using an RCT evaluation design to determine if and when payments would be made.

Both PFS grantees reported they developed tightly constructed contracts to formalize all aspects of the pilot partnerships and to ensure that their perspectives and interests were addressed when the initiative was operationalized. The key topics addressed in each state’s partnership contract were defining the intermediary’s and other partners’ responsibilities; documenting the entire evaluation plan including the specification of the RCT design and related methodologies and expected outcomes of the service intervention; and savings that could potentially be realized. The contract also specified

the roles of the evaluator, the independent validator and the process for determining how payouts to investors would be handled. In addition, the contracts reflected that the organizational relationships needed to extend beyond the traditionally established government appropriation and contracting cycles and that the pilots needed to be implemented under the aggressive timeline specified by the DOL SGA.

The PFS partners reported that each of these issues required deliberation and negotiation, including detailed planning and articulation of “what if” scenarios to identify possible unanticipated circumstances. The result of these negotiations was a detailed partnership contract drawn up by each state’s legal team. Topics that required particular attention included client referral processes, determining investment risk and returns for investors, and the coordination of the different state data systems needed to generate outcome information. The PFS partners reported that the contract negotiations were painstaking but necessary. In addition, the process was constrained by the DOL SGA’s deadlines, which motivated the partners to reach agreement on details. While not formal signatories to the contract, the investors in each pilot are represented indirectly by the intermediary who negotiated their initial engagement and is responsible for involving them in relevant design, operational and financial issues.

The design and operation of the PFS pilots also required a management structure that allowed for ongoing communication, transparency, and timely decision making, both initially and over the project life cycle. The management structure for the partnerships was established as part of each state’s contract, both of which specify a committee structure, with committees responsible for strategic direction and operational oversight. The contracts specify the management or oversight of committee composition, structure, attendance requirements and flexibility, meeting frequency, and specifications for determining meeting agendas. The New York contract also specifies the exact roles and responsibilities of each committee, specific individuals involved, and an explicit management process for altering the oversight structure or processes. The PFS partners reported that these specified committee and meeting structures were important for managing the partnerships, particularly given that PFS was a multi-year effort. The contract agreements and management structure for each of the pilots are summarized in detail in the interim report.

### **2.1.4 Raising Capital and Identifying Investors**

The PFS model relies on investors to finance a public service intervention in exchange for a potential return on their investment. This section describes how each pilot recruited investors and the financial investments they made.

The PFS intermediary in each state played the lead role in recruiting investors. The intermediaries were selected in part because of their understanding of the pilot requirements and their ties to private capital markets and philanthropic sectors. While the prospective payment of investors was funded from two distinct sources (DOL in Phase I and state resources in Phase II), the intermediaries solicited the PFS opportunity to investors as a single initiative, that is, investors and their capital were not specifically earmarked for the different phases of the pilot.

Both pilots sought a mix of private sector and philanthropic-based investments to increase the type and number of project stakeholders. Staff from both pilots reported that the involvement of philanthropic partners increased the credibility of the investment opportunity among private sector investors. As part of investor recruitment, the staff from PFS pilots reported they had to educate

## THE DESIGN AND IMPLEMENTATION OF THE PFS PILOTS

investors about the PFS concept and the RCT evaluation design that would be used to determine whether payments could be made. Another key issue for investors was whether the government funds intended to pay back investors if outcomes were met would be available, given the constraints of annual appropriations.

Massachusetts and New York recruited investors to participate in the pilots but used a different financing model in the investments, with Massachusetts relying primarily on a loan from a single investor (Goldman Sachs) and New York using one financial investor (Bank of America Merrill Lynch) to secure investments from numerous investors through a Private Placement Offering (PPO). Exhibit 2-3 provides an overview of the financing structure in each pilot and how the financing mechanisms operated.

### Exhibit 2-3: PFS Financing Approach for Each Pilot

Financing Features		Massachusetts	New York
Led Investor Recruitment		Third Sector/New Profit Inc.	Social Finance/Bank of America Merrill Lynch
Capital Raised		\$19.4 million	\$13.5 million
Type of Financing		Working Capital Loan	Securities
Description		Goldman Sachs financed senior loan through its Social Impact Fund (a diversified investment fund of socially conscious investments) Living Cities & Kresge Foundation provided junior loan financing	Bank of America Merrill Lynch <sup>a</sup> served as the placement agent offering the securities through a PPO to high-net-worth accredited investors
Investors	Debt <sup>b</sup>	<ul style="list-style-type: none"> <li>Goldman Sachs</li> <li>Kresge Foundation</li> <li>Living Cities</li> </ul> Deferred fees from Roca & Third Sector	<ul style="list-style-type: none"> <li>40+ high-net-worth accredited investors with at least \$10M invested within Merrill Lynch</li> <li>Laura &amp; John Arnold Foundation</li> <li>Living Cities</li> <li>Kresge Foundation</li> </ul>
	Grant	<ul style="list-style-type: none"> <li>Laura &amp; John Arnold Foundation</li> <li>New Profit Inc.</li> <li>Boston Foundation</li> </ul>	<ul style="list-style-type: none"> <li>Robin Hood Foundation</li> <li>Rockefeller Foundation (loan guarantee)</li> </ul>
Conditions		Investors are made public; they are involved in monitoring pilot. Early termination clause gives investors opportunity to exit pilot.	Investors are confidential. The investment adheres to SEC regulations. Not flexible or transferable. No easy exit strategies or automatic triggers if the pilot is not performing.
Repayment		Repayment of the principal is on a prioritized basis, with the senior loan repaid first followed by the junior loan and then Roca and Third Sector's service fees	Repayments are divided up equally among the private investors based on level of investment.

<sup>a</sup> BAML did not invest any of its own money, instead their contribution was developing the PPO and marketing it to their clients.

<sup>b</sup> Investors listed under "Debt" include organizations and individuals who are eligible to receive a payment if outcomes are achieved.

---

## THE DESIGN AND IMPLEMENTATION OF THE PFS PILOTS

---

The DOL SGA required applicants to have the expected investor capital preliminarily secured at the time grant applications were submitted in January 2013, and the PFS applications had to include investor letters of intent/interest. Most investors were still finalizing the terms of their commitments in spring 2013, and the details of the financial structures (outcome payments and timing) were not finalized until fall/winter 2013. The next sections describe in more detail the involvement of the investors in each state in more detail.

### Massachusetts

While Third Sector led the effort to recruit investors in Massachusetts, New Profit, Inc., a philanthropic organization with a focus on promoting social innovation and building capacity in nonprofit organizations, also played an important role. New Profit had established credibility in the juvenile justice arena as well as contacts in the private investment and foundation communities, and the pilot partners reported New Profit's involvement facilitated the identification of other investors. New Profit also made an investment of its own (see below).

The intermediaries' staff reported an important element of their outreach was familiarizing potential investors with the PFS approach, as many did not have experience with the proposed financing and payment structures. The one exception was Goldman Sachs, which had previous experience with the PFS model through its investment in the New York City Project at Rikers Island. During the PFS grant application period, Third Sector and New Profit secured \$19.4 million of aggregate commitments of which \$11.7 million are required to fund project operations under Phase I. The primary investors' contributions are summarized below:

**Goldman Sachs.** This investment firm agreed to provide \$8 million in loans to the PFS pilot and is the project's "senior lender," meaning its investment will be paid off first if and when there are returns. To develop the agreement, Third Sector worked with Goldman Sachs's Urban Investment Group, which has a focus on partnering with community stakeholders to support underserved urban communities.

**Living Cities and the Kresge Foundation.** These philanthropic organizations each agreed to provide \$1.33 million in loans for the PFS pilot. Both joined as "junior lenders," meaning they would be repaid after the senior lender, Goldman Sachs, was paid in full. Living Cities is a consortium of 22 foundations and financial institutions (including the Kresge Foundation) that seek investment opportunities to address the social and economic challenges facing low-income individuals. At the time Massachusetts was seeking investors for its pilot, a number of Living Cities' member organizations were trying to advance the PFS approach within the consortium as part of their broader emphasis on performance-based investments. Beyond its contribution through Living Cities, the Kresge Foundation made an additional contribution due to an established history the foundation had with Roca, the pilot's service provider.

**Laura and John Arnold Foundation, New Profit, Inc., and the Boston Foundation.** These three foundations provided "recyclable grants" to the PFS pilot. This means that the foundations made the grants to the PFS initiative with no expectation of a return. Instead, any returns on their investments are to be "recycled" into other PFS initiatives in Massachusetts or to expand Roca into new geographic areas. Recyclable grants came from the Laura and John Arnold Foundation (\$3.34 million), the Boston-based New Profit (discussed above, \$1.81 million), and the Boston Foundation (\$300,000). The grant from the John and Linda Arnold Foundation, which has a national reputation

for supporting evidence-based approaches to address social problems, was viewed as significant by others considering an investment as it allayed their concerns that the initiative had only local support.

**Third Sector and Roca.** In addition to these loans and grants for the pilot, both Third Sector and Roca deferred payment of a portion of their program costs to provide additional financial support to the pilot. Under this arrangement, a portion of Third Sector's intermediary costs and Roca's service provider costs are being withheld until certain benchmarks are achieved. The investors interviewed for this study reported that these deferred payments were another important factor in their decision to participate because it showed Third Sector and Roca had confidence in their own services and performance.

The Massachusetts PFS pilot operates through an expansion loan from the private sector investors. This means that the money to operate the Roca services and manage the PFS pilot is loaned by investors to Third Sector/Youth Services Inc. and Roca. Pilot partners reported that they used this financing approach because it was preferred by Goldman Sachs, the pilot's largest investor. Goldman Sachs's Massachusetts PFS investment was one of the first investments it made as part of its \$135 million Social Impact Fund and was a key element of its broader business plan to support a diversified array of socially conscious investments. Goldman Sachs preferred the straightforward expansion loan structure for managing the financing because the structure could readily be replicated for future investments.

At Goldman Sachs, the investment opportunity was reviewed by an internal investment committee, which is standard practice at the company. The investment committee examines all risks against potential returns. In its deliberations about the project, the committee examined Roca's service model; its' potential to achieve the performance measures, or project outcome targets, that would result in repayment of the loan; and its overall organizational capacity, including managerial experience, staff stability, and financial solvency. Goldman's investment committee also examined the risks related to the state legislature's commitment of \$15.3 million to the project. The federal government shutdown of 2013, which occurred after the grants were awarded but while the investment was being finalized, also raised concerns. Goldman staff reported that the governor's establishment of the Social Innovation Financing Trust Fund, which provides up to \$50 million for PFS projects, helped to ease Goldman's concerns.

In its capacity as junior loan partners, Living Cities and the Kresge Foundation also reported that they assessed the risk of the investment before committing to the pilot. While acknowledging the assessment done by Goldman Sachs, the foundations hired an outside expert to examine the potential risks and rewards from their organizational perspectives.

### **New York**

The New York intermediary, Social Finance, was responsible for recruiting investors and was selected because of its knowledge of the PFS model. Social Finance approached several financial institutions and philanthropic organizations to identify potential investors and financing options. Ultimately, Social Finance and the other partners selected Bank of America Merrill Lynch (BAML) to develop the PFS financing for the pilot while also approaching several philanthropic organizations. The investors and the financing approach for the New York pilot are described below.

**Bank of America Merrill Lynch.** BAML is the New York pilot's placement agent working together with Social Finance to raise the full capital amount. The approach to raising the capital was to



## THE DESIGN AND IMPLEMENTATION OF THE PFS PILOTS

---

establish a private placement offering (PPO), a non-public security offering made available to a limited number of pre-screened investors identified by BAML.

Social Finance reported that it chose to work exclusively with BAML for two reasons. First, the investment opportunity suited BAML's established business priorities. BAML staff reported having identified strong interest among groups of investors for additional "impact investing" opportunities, the term used for investments that target social problems, and PFS met this objective. Second, staff at BAML were knowledgeable about the New York City Project at Rikers Island, which reduced their perception of the risks associated with the approach.

BAML's preferred approach to accumulating the capital necessary for the PFS pilot was to establish a PPO. Staff at BAML reported that their clients prefer PPOs because of their relative simplicity as compared to other investment mechanisms, with investment returns tied directly to performance and payments divided proportionally based on the level of investment. The PFS PPO fell under BAML's Wealth Management Division, which offered the investment opportunity to specific high-net-worth investors (those with at least \$10 million invested with BAML, among other requirements). BAML staff reported that these investors were among those with an interest in social impact investment opportunities and thus were good candidates for investing in the PFS pilot.

BAML and Social Finance worked collaboratively to develop the PPO, which required following standard BAML procedures for assessing the opportunities and risks of the investment. This necessitated obtaining incarceration data from DOCCS and program performance and cost data from CEO, and using the data to identify acceptable performance thresholds, payment triggers, and return potentials.

These analyses also included assessing risks that could potentially impede the achievement of the goals that would trigger payouts. BAML staff reported they had concerns about whether the New York State legislature would maintain its funding commitment. To address these concerns, the partners agreed to an accelerated evaluation design for reporting results and determining payment only in the case of an appropriation failure resulting in early termination of the project. Specifically, if the state legislature was not able to allocate enough money to fully repay investors for the outcomes achieved, the PFS pilot would repay investors proportionally based on the money available. In addition, concerns were raised about incentives for the intermediary and service provider to drop out of the pilot if the project was underperforming and appeared unlikely to meet the established outcomes. In response, the partners specified circumstances or conditions that would warrant replacement of either the intermediary or the service provider (e.g., non-responsibility, confidentiality breach, bankruptcy). Finally, as in Massachusetts, the sustained availability of DOL grant funding was considered an area of concern, given that the 2013 federal government shutdown was being debated as the investors were finalizing their investments.

Once these issues were resolved, BAML created a PPO security called the Social Finance N.Y. State Workforce Reentry 2013 LLC, which was marketed to BAML's investors, with \$100,000 established as the minimum investment. Typically, the outreach and recruitment of individual investors for a PPO at BAML is conducted by an internal team of financial advisors. For the PFS project, however, the unique nature of the investment, coupled with a compressed time frame (recruitment of individual investors had to be accomplished in approximately half the time that BAML generally uses to market a PPO), Social Finance joined BAML and played a key role in marketing the investment.

To recruit investors, Social Finance and BAML developed materials that provided an overview of the PFS pilot; explained the contract between CEO and Social Finance; described the policy issue and target population, particularly recidivism and its effects; and explained the investment opportunity and how the financial vehicle would work. Since a PPO is a federally regulated financial vehicle, Social Finance and BAML had to abide by Securities and Exchange Commission disclosure rules, meaning the risks and the potential for a complete loss of the investment had to be disclosed as part of its recruitment efforts.

In addition to individual private investors, Social Finance identified and recruited foundations to invest in the PFS project, but as described below these investors had different expectations about the payoffs they wanted from their investments.

**Robin Hood, Laura and John Arnold Foundation, the Kresge Foundation and Living Cities.**

The Laura and John Arnold Foundation provided \$4 million and the Robin Hood Foundation, whose mission focuses on poverty reduction, provided \$300,000. These foundations specified that any returns on their investments were to be used to support future social innovation projects, with the goal of rigorously evaluating programs and scaling up those that are proven to have an impact. Finally, the Kresge Foundation and Living Cities, invested \$100,000 and \$509,000 respectively with particular interest in comparing the experience of being an investor in the two PFS pilots.

**The Rockefeller Foundation** provided \$1.3 million as a 10 percent “first-loss guarantee” for all investors, reducing these investors’ risk to 90 percent of their investments if the performance targets are not met.

Staff at BAML and other PFS partners reported that the philanthropic involvement was important for securing private sector investments. Specifically, the Robin Hood Foundation’s knowledge of the target population and intervention strategies and the Laura and John Arnold Foundation’s interest in evidence-based investment brought credibility to the project.

## 2.2 Developing the PFS Pilots’ Evaluation Designs

Evaluating the success of the service intervention in achieving its objectives is a critical component of any PFS project, and this section describes the evaluation designs implemented for each pilot. Determining whether the intervention had the desired impacts is not only important in terms of understanding if and how much investors will be repaid but also for determining the benefits to participants and communities. Specifically, the evaluations of the PFS pilots were designed to provide rigorous evidence on whether the service interventions were effective in increasing employment and reducing recidivism.

Recognizing the important role of evaluation in a PFS project, DOL’s SGA required applicants to use an RCT or credible quasi-experimental design, and both pilots used an experimental design in their projects. This requirement was included to help ensure that the pilots’ investors would only receive payouts for results that could be causally linked to the services provided.

An RCT requires eligible program applicants to be assigned at random (a lottery) to either a group that can access the tested service intervention (the treatment group) or a group that cannot access the intervention but can access other services in the community (the control group). For the two PFS pilots, this means control group members cannot be referred directly to the intervention under study at

the PFS service providers, although they can seek out services at these or other organizations on their own.<sup>9</sup> Because of the random assignment, the two groups are identical at the time of entry into the intervention and any differences in outcomes detected between them during the follow-up period can be attributed to the service intervention. The difference in outcomes between the treatment and control group is referred to as the “impact” of the service intervention under evaluation.

Massachusetts chose to engage a third party to operationalize and oversee its evaluation plan. Sibalytics was initially chosen as the evaluator for the PFS pilot in Massachusetts but was replaced ten months after the pilot started with the Urban Institute. Alternatively, New York determined that the Research Office of the state Department of Corrections and Community Service (DOCCS) would be responsible for all aspects of the evaluation including the random assignment of the sample and securing secondary data for measuring individual outcomes.

Below we highlight several key elements of the evaluation plans including each pilots’ approach to identifying and selecting their sample and their choice of data sources for measuring approved outcomes. This is followed by a discussion of the pilot’s choice of back up methodologies in the event that the RCT proved impractical or infeasible. The section closes with an overview of the role of the independent validator that each grantee engaged to review the methods and findings of the evaluators. Further details on each pilot’s evaluation design are provided in Chapter 4.

### 2.2.1 Specifying Evaluation Designs, Sample Sizes, and Data Collection

While the PFS SGA allowed for some flexibility, both states generally recognized DOL’s interest in implementing an RCT evaluation design whenever feasible. However, many of the partners reported having a limited technical understanding of this type of design and what it would require when they began developing the pilot projects. Technical assistance provided by the Harvard GPL helped the pilot partners understand the concepts involved and develop design options. While some partners initially were concerned about implementing an RCT design due to costs and operational requirements, partners from both pilots reported that the Harvard GPL’s assistance helped address these concerns and launch the designs. In addition, a number of evaluation design issues had to be addressed as part of the planning process, such as setting the enrollment period for the study so that outcomes could be estimated in the time frame necessary to pay back investors and determining the sample sizes needed to detect desired impacts. As a result, developing the details of the evaluation design was a significant activity for both pilots. The key issues in developing the evaluation are described below.

**Determining sample sizes and enrollment period.** In developing their evaluation designs, both pilots had to determine the sample size needed to detect the impacts that would trigger payments. In general, random assignment designs need sufficient numbers of individuals assigned to the treatment and control groups (sample size) to detect statistically significant impacts. This is especially true for designs where not all treatment group members participate in the service intervention after they are randomly assigned and referred to it by the probation staff. Some possible participants may choose not to participate, which was the case for the PFS pilots. That is, the intervention can only reasonably

---

<sup>9</sup> See discussion in Section 2.2.1 for how the evaluation design accounts for the issue that the control group members may receive similar services. See also Appendix B for a general definition of the RCT evaluation design.

affect those who receive at least some of its services. Based on calculations they conducted, Harvard GPL staff determined that the pilots designed in response to the DOL SGA were “underpowered”; that is, they lacked sufficient sample size to detect the required impacts at desirable levels of statistical confidence.

As discussed above, this conclusion led both pilots to expand the DOL pilot to a second phase supported by state funds. While the parameters of the pilot proposed in the grant application to DOL (Phase I) would remain intact, the addition of a second phase would allow for increasing the sample size necessary to detect precise estimates of impacts with acceptable levels of precision that trigger payments to investors. This strengthened the overall value of the pilots by increasing their capacity to generate budgetary savings to government.

As set out in the application to DOL, Massachusetts aimed for a target number for enrollment in Roca services at 535 for Phase I and 929 for Phase I and Phase II combined. To achieve the enrollment targets in its PFS proposal to DOL, the Massachusetts pilot aimed to randomly assign 753 to the treatment group. The partners for New York State pilot set the target for the number enrolled at CEO at 1,000 for Phase I and 2,000 for Phase I and Phase II combined.

**Identifying data sources needed for the evaluation.** Another issue that both pilots addressed was identifying sources for the data needed to conduct the evaluation. These data needed to include the demographic characteristics of all individuals at the time of random assignment (known as baseline data), the duration and content of services delivered to treatment group members by the two service providers, and the employment and recidivism outcomes for both treatment and control group members needed to calculate investor payments.

Collecting the data needed for the evaluation required the involvement and coordination of several different agencies in each state. As part of the planning process, each of the pilots determined the parties responsible for providing the data and ensuring that it would be available on the schedule needed to make payouts to investors. In Massachusetts, a number of agencies and organizations were involved in collecting baseline data, while recidivism data were collected from the Executive Office of Public Safety and Security’s Criminal Offender Record Information Support Services Unit. Employment data, specifically Unemployment Insurance (UI) quarterly wage records, come from the Executive Office of Labor and Workforce Development. In New York, the Department of Corrections provides baseline and recidivism data and NYSDOL provides the employment data (again UI wage records). For both pilots, the service providers provide data on service receipt.

**Identifying appropriate statistical methods to compute impacts.** The pilot partners also worked with the Harvard GPL to determine the appropriate statistical methods to use in the evaluation. Because some treatment group members may not choose to participate in the PFS interventions and some control group members may access these services on their own, the pilot partners worked to develop statistical methodologies that reflect that not all treatment group members will receive the services being studied, while some control group members may receive some type of services.

**Determining a back-up evaluation strategy.** Each pilot opted to have a back-up methodology planned in the event that their proposed RCT was not viable, in case there was not a large enough sample overall (if enrollment proceeded at a slower than anticipated pace and not enough participants enrolled in the specified time frame) or if too many members of the control group accessed the services, which would lead to insufficient “contrast” in outcomes, between the treatment and control

group members who received services. As discussed, such a possibility might occur, since control group members could enroll in the program on their own rather than through a referral generated under random assignment procedures.

In Massachusetts, the specific trigger for implementing the back-stop methodology was if the proportion of treatment group members served by Roca was not at least 30 percentage points higher than the proportion of control group members that received Roca services. The Massachusetts partners developed a non-experimental difference-in-differences analysis that would compare changes in outcomes for high-risk individuals in the communities served by the service provider (Roca) that are included in the PFS pilot to changes in the outcomes of similar individuals from other communities not being served by Roca. The Massachusetts pilot planned to use the back-stop methodology (if needed) at the end of Phase II, which is the end of the state-funded project. Phase I used the RCT evaluation design because the difference between the percentage of individuals in the treatment group receiving services and the percentage of individuals in the control group receiving services were greater than 30 percent.

In New York, the trigger for implementing the back-up evaluation strategy was the same as in Massachusetts—the proportion of treatment group members receiving from CEO must be at least 30 percentage points higher than the proportion of control group members receiving these services from CEO. The New York PFS pilot team developed a non-experimental matched historical comparison evaluation as its back-up methodology. Under this approach, individuals in the treatment group would be matched to similar individuals, based on their baseline characteristics, who were released from prison before the start of the PFS pilot in New York City and Rochester. In the end, the RCT was used for Phase I of the project.

**Payment formulas for investors.** Both pilots also developed formulas for calculating payments to investors, based on the impacts achieved. While the specifics were different, both pilots had thresholds or tiers, which involved an increasing level of payment per capita at the different thresholds, for both the employment and recidivism impacts, multiplied by the number of participants.

**Defining the role of the independent validator.** Each pilot retained an external independent validator to oversee the implementation of the evaluation, monitor its fidelity and validate the findings. In addition, the validator was responsible for confirming that data are being interpreted correctly, that outliers and missing values are being handled properly and that outcomes are statistically unbiased. Once the results of the evaluation are determined, the validator reviews the data and evaluation methodology to verify the results that trigger incremental success payments. As part of planning the grant application, pilot partners worked to select a validator and determine how its work would be coordinated with the evaluation. During the early stages of random assignment and enrollment, the independent validators worked closely with the evaluators to monitor referrals and ensure random assignment followed the evaluation plans. The independent validators were Public Consulting Group (Massachusetts) and Chesapeake Research Associates/Mike Puma Associates (New York).

### 2.3 Summary

This chapter describes the circumstances that led to the development of the two pilots, their general structure (including partners and contractual arrangements), and the development of their evaluation

## THE DESIGN AND IMPLEMENTATION OF THE PFS PILOTS

---

designs. In both states, the pilots built on existing momentum to establish PFS approaches and pro-bono support from entities supportive of the approach, such as the Harvard GPL. The DOL solicitation accelerated that momentum by requiring the states to formalize partnerships within a specific timeframe. Both pilots successfully engaged all of the required partners, including the government agency, intermediary, investors, and validator, but other partners were also engaged in key roles, including the service providers and technical assistance providers in both states, and an external evaluator in Massachusetts. The pilot partners developed extensive partnership agreements and comprehensive structures for managing the projects. In both states, partners were able to raise the necessary capital and reported that philanthropic support was beneficial for obtaining support from private sector investors.

The DOL SGA specified that the pilots had to use a rigorous quasi-experimental or experimental research design, and both pilots opted for the RCT design. Some pilot partners were concerned about the cost and operational complexity of the evaluation design, and the Harvard GPL helped address these concerns. Additional evaluation design issues included identifying data sources needed for the evaluation, identifying appropriate statistical methods, developing a back-up methodology and defining the role of the independent validator.

### 3. PFS Pilot Interventions

This chapter describes the services provided through the two PFS pilot interventions by the Center for Employment Opportunity (CEO) in New York, and Roca in Massachusetts. The Abt team collected information about the interventions through interviews with program staff, a review of documents and informal, small focus group discussions with program participants.

For each of the pilots, the chapter describes the target population, the theory of change, and the key components of the intervention’s services. While neither service provider made changes to their core service model under PFS, they did seek to some changes under the grants. The PFS approach requires a balance between using a program model that is considered promising enough for investors to feel confident that it is likely to achieve outcomes targets, but to be enough of a departure from “business as usual” that the government agency would be unlikely to pay for services outright. In both states, partners elected to utilize a long-standing model but leveraged the PFS funding to be provided by DOL to increase their reach—in Massachusetts by expanding Roca’s service area, and in New York by proactively targeting recent parolees.

#### 3.1 The Intervention in Massachusetts: Roca and its Service Model

This section describes the program model and services provided by Roca under the PFS pilot in Massachusetts, including changes the organization made to its interventions under the pilot, and the experiences in the different locations where the intervention was made available.

Founded and established in 1988, Roca aimed to break cycles of poverty and incarceration by engaging youth in positive, intensive relationships, which lead to behavior change, the development of life skills, and participation in education and employment. The intervention was guided by a model of behavior change called “Stages of Change” which posited that sustained behavior change involves movement through the following five stages: pre-contemplation, contemplation, planning, action, and sustaining. During the “pre-contemplation” stage, most participants are not yet ready, willing, or able to make a change. As they moved through the stages, participants were to examine their behaviors and consider changing them, prepare to change, begin to behave differently, and then maintain that change. Relapse, or backward movement through these stages, can happen at any point in the process.

Roca participants were primarily males under age 25. Roca, or “rock” in Spanish, served a range of disadvantaged youth, including high school dropouts, gang members, young parents, refugees, and immigrants and specifically targeted individuals with a history of involvement in the criminal justice system. Virtually all participants have been involved in the juvenile and/or adult criminal justice

**Key Features of the PFS Interventions**

- *Both service providers had long track records and solid reputations in their field, and the capacity to manage data and participate in an RCT.*
- *Both interventions targeted individuals involved with the criminal justice system.*
- *Both service providers changed recruitment procedures under PFS and relied primarily on referrals from outside agencies to facilitate random assignment and meet enrollment targets.*
- *Under the PFS pilots, Roca expanded the program to new sites and made changes to recruitment and enrollment procedures.*
- *CEO changes under the PFS pilot were also primarily adjustments to the recruitment, referral, and enrollment processes.*

system and had limited or no work histories and minimal educational attainment. Because of their history, most participants are unable or unwilling to enroll in standard employment and training programs available in the community.

The program was to facilitate behavior change through “cognitive restructuring.” This was a therapeutic process in which individuals learned to identify, understand, and change negative thought patterns that can occur in stressful situations through work with trained counselors. Staff reported that once established, participants’ relationships with staff became the foundation for cognitive-behavioral change as they help participants develop a plan for improving their lives and creating opportunities as they make educational, life-skills, and employment gains. Roca used a cognitive behavioral therapy curriculum developed in partnership with Massachusetts General Hospital which focused on ten skills that helped individuals develop emotional regulation, acceptance, and problem solving. The Roca model recognized that the process of behavior change can take years and included two years of intensive services and two years of follow-up services.

### 3.1.1 Services Provided by the Roca Program

Overall, the Roca program involved two years of engagement in services that were customized to participant needs, and two years of follow-up services. Services included a risk assessment; one-on-one assistance from program staff; a variety of education and employment activities developed specifically for high-risk, underserved populations; coordination with institutions in the community; and the use of data to inform program management. Specifically, the services included:

**Relentless Outreach.** Roca’s outreach methods, prior the PFS pilot, involving finding participants through an approach called “relentless outreach,” which consisted of program staff seeking to contact eligible youth by going to “where they are.” This could mean connecting with youth directly in the community (e.g., circling the neighborhoods in the Roca van or on foot, knocking on doors) or within the criminal justice system as individuals prepare for release from jail, court, and sometimes from referrals from the Department of Youth services or parole officers.

**Transformational Relationships.** The core component of the Roca program was strong personal relationship between participants and program staff. Underlying this approach was the theory that strong relationships with trained counselors can reinforce the skill-building and behavior changes needed for participants to establish new patterns of decision making. Once individuals enrolled in the program, they were assigned to a youth worker who followed the participant throughout their tenure in program. Youth workers were trained on a range of evidence-based and clinical techniques for promoting behavioral change, including cognitive restructuring, client-centered counseling, and “stage-based” learning that focuses on academic and prevocational training and transitional employment.

**Assessment and Sustained Engagement.** Youth workers first conducted enrollment and an assessment of risk, including a substance abuse screen, relationship screen, and a life distress screen. As part of the enrollment process, individuals completed a central intake form that collected information about participants’ demographic characteristics, receipt of public benefits, housing status, and employment. Also recorded on this form were the results of assessments, and the form was updated every six months through a “change assessment.” Youth workers carried caseloads of no more than 25 participants and aimed to communicate with each participant in their caseload a



minimum of twice per week. Youth workers aimed to hold these meetings with participants in-person, but would conduct meetings by phone if necessary.

**Individualized Employment Services.** For the next step in the program, the youth worker and participant jointly developed an individualized service plan that articulated goals and the services available to help the participant meet his goals, in particular the goals of finding employment and staying out of prison. The individualized assistance and support provided by Roca youth workers drew from a range of employment-focused activities provided on site by staff. The activities used and their sequence are tailored to individual participant’s needs and circumstances and included the following:

- **Education and training activities.** Roca provided a wide range of education and training activities. In addition to English as a Second Language (ESL) and GED classes, Roca provided on-site training courses in such skill areas as forklift operation, green cleaning and housekeeping, construction, hospitality, retail, and culinary arts. Roca also offered training for industry-recognized certifications in such areas as cardiopulmonary resuscitation, automated external defibrillator use, safe lead removal, carpet care, floor care, and food handling safety.
- **Work-readiness activities.** Roca offered workshops on employment-related topics such as workplace communication, resume development, job interview skills, conflict resolution, and sexual harassment sensitivity. Activities also included an emphasis on developing the “soft skills” needed for employment, such as being on time, working on a team, following directions, having a positive attitude, and exhibiting appropriate job behavior.
- **Transitional employment.** Transitional employment was subsidized employment, with all or part of the wage funded by Roca, which aimed to teach individuals the job skills they needed to transition to unsubsidized employment. To help participants learn workplace behaviors and gain work experience, participants were sometimes assigned to a small work crew or placed individually in a job based on their interests and training. Most participants were encouraged to participate in this component. Work crews, supervised by Roca staff, consisted of seven to eight participants working on property maintenance, painting, or landscaping. Advanced work crews contained participants who were not yet ready for job placement but were capable of taking on more responsibility on the job.
- **Job placement.** Job developers tracked employment opportunities in the area and communicated with local employers about appropriate job opportunities for their clientele. Staff also worked directly with participants to help them find unsubsidized jobs.
- **Post-employment follow-up.** Roca continued to provide follow-up career guidance and support services to individuals once they found unsubsidized employment for up to two years.

**Phased Services.** Within the context of sustained engagement, the Roca program was intended to be flexible enough to accommodate individual circumstances. Consequently, the participants moved through the program in phases, with time spent in each phase depending on individual needs and circumstances.

- **Phase 1.** Participants met with youth workers on a regular basis, preferably at least twice weekly. Roca staff worked on building trust with participants, and they monitored participants’ progression through Roca’s “Stages of Change.” Youth workers expected some participants to

relapse, or move backwards through the stages, before they course-corrected and moved forward and on to Phase II. Participants moved into Transitional Employment after they have made at least ten contacts with their youth worker, which typically took about two months. If a youth worker had made contact with an individual while they were still in prison, this counted toward the ten contacts.

- **Phase 2.** Participants continued to meet with youth workers and engage in program services, including transitional employment. In order to begin receiving job placement services, participants had to have been employed on a work crew for 60 consecutive days. Participants were removed from work crews after they receive four “write-ups” for poor behavior, and it typically took a participant 20-30 days to get back on a work crew. On average, participants were removed from work crews six to seven times, and it typically took individuals 15 to 18 months to meet the 60-day requirement.
- **Phase 3.** The frequency and types of services changed in the last two years of the program, known as Phase 3. During this time, participants were assigned to a new youth worker with a larger caseload, and they tended to meet with that person less frequently. Youth workers assigned participants to one of three “levels” of need during this phase. Those in “Level 1” were working and generally considered stable. These participants meet with their youth workers twice a month. Participants in “Level 2” were working but were not stable and met with youth workers more than twice a month. Participants in “Level 3” were not hitting their benchmarks, often due to substance abuse or other issues.

**Performance-Based Management.** Roca used an internal program management system to track the program participation and monitor participant progress through the phases. Roca staff used this system for PFS (specifically for the job readiness outcome) as well as for other contracts. Once individuals were enrolled in the program, staff tracked their contacts with participants, including the number of times they make contact, the duration of the contact, and its outcome. The system also tracked the types of services in which participants engaged and more detailed information about their participation in transitional employment (e.g., number of hours, warnings, termination). Roca used these data to assess each participant’s level of engagement in Roca, from “not engaged” to “consistently engaged.”

### 3.1.2 Changes in the Roca Program under the PFS Pilot

Roca staff reported few changes to the services provided as a result of the PFS program, but as described below, PFS funding allowed the program to increase in scale in several ways.

**Program expansion to new locations and service areas.** The PFS funding allowed Roca to serve to more participants by expanding catchment areas from original locations in Chelsea, East Boston, and Revere to new locations north of Boston (Cambridge, Charlestown, Everett, Lynn, Malden, Medford, Somerville, and Winthrop). Roca also added a site in Lynn, which operated as a satellite to the Chelsea site. The PFS pilot also allowed Roca to open new sites in Boston and Springfield. These sites and the changes under PFS are described below.

- **Boston.** Under the PFS pilot, to increase program enrollment, Roca opened its first site in Boston. Roca staff reported that the majority of participants at this location are affiliated with one of the 200 gangs located in the city, and as a group, the participants have a high-risk level according to

Roca’s assessments. This resulted in some changes to the “standard” Roca services. Due in part to the high rates of gang affiliation as well as an incident of violence that occurred at the site early on in the PFS pilot, the Boston Roca office is a “closed” building with locked doors, metal detectors, and restricted access for all participants. As a result, Roca staff created a schedule of programming each week that ensured that participants in rival gangs are not in the building at the same time. Because of the need to consider gang affiliation in work crew assignments, staff reported this limited flexibility in work crew scheduling and that high risk participants were often placed in job placement programs instead. Restrictions on access to the building and work crews also reduced flexibility in scheduling; if participants could not attend their assigned time slot in a given week, they may have experienced delays in being reassigned to a new work slot. The “closed” building also did not provide participants a safe, caring place to go, which was viewed as a key element of the program.

- **Chelsea.** Under the PFS pilots, Roca expanded the service area for Chelsea to include areas north of and in Boston and in Springfield. The Chelsea office is the original Roca location and, as a result, had long-time, established relationships with community partners. Staff reported that many participants in Chelsea struggled with substance abuse and youth workers dedicated significant time to addressing this issue.
- **Lynn.** The Lynn office opened as a satellite to the Chelsea location under the PFS pilot. This site has been hit hard by the opioid epidemic; staff reported that about two-thirds of participants in this site actively abuse at least one subset of drugs. In response, one staff was dedicated to enrolling individuals in substance abuse treatment programs. Roca staff also reported that the Lynn office provided referrals to behavioral health services more than the other Roca locations.
- **Springfield.** This office also opened to expand Roca enrollment, and staff reported that the population served had significant gang activity. However, this did not necessitate as many changes to Roca’s programming as in Boston since the gangs were geographically spread out, and the site is able to maintain an “open” building and participants are better able to attend work crews.

**Increased use of referrals from organizational partners rather than direct participant recruitment.** Roca staff reported that the most notable change in their program under the PFS pilot was the method of recruitment. As noted above, prior to PFS, Roca staff recruited participants by going to “where they are,” whether connecting with youth directly in the community or within the criminal justice system as an individual prepares for release. Under the PFS pilots, Roca received referrals of youth and young adult men ages 17-24, who were on adult probation or aging out of foster care, and living in one the Roca locations. As noted later in the report, the referrals originated in several state and county agencies, including the Office of the Commissioner of Probation (OCP), the Department of Youth Services (DYS), the Massachusetts Department of Corrections (DOC), the Massachusetts Parole Board (MPB), and sheriff’s departments from program locations.

### 3.2 PFS Pilot Intervention in New York

The service provider for the New York PFS pilot was CEO, a nonprofit organization that provided employment-focused services for men and women returning to their communities after being released from incarceration. CEO targeted individuals who were hard to employ, primarily those with criminal

backgrounds and minimal employment experience. CEO's headquarters are in New York City, but it operates at over a dozen sites in seven states (New York, California, Colorado, Ohio, Oklahoma, Pennsylvania, and Tennessee).

The focus of the CEO model, which was used across all of its locations, was to engage participants in transitional employment soon after their release from incarceration to provide earnings, avoid parole violations related to unemployment, and build workplace skills. Similar to Roca in Massachusetts, CEO has a long history of operating this program model (the organization was founded in 1996). With the DOL grant, CEO planned to serve 1,000 individuals in New York City and Rochester, New York, targeting formerly incarcerated individuals under community supervision who were at high risk of re-incarceration.

CEO was previously the subject of a RCT, conducted by MDRC, whose report was published in 2012. That evaluation found that the program had initially increased employment due to the transitional jobs, but the employment effects faded over time. However, the program also significantly reduced recidivism, with the greatest impacts among those who enrolled shortly after release from prison, and for those who had been assessed by the Corrections Department to be as at high risk of recidivism (Redcross et al. 2012, 27)

### 3.2.1 CEO Program Model

The CEO program involved a sequence of activities, including:

**Enrollment.** Prior to the PFS project, potential participants typically learned about CEO through the parole office where CEO was advertised as an employment agency that could support parolees in securing employment. Some prospective participants also learned about the program from family or friends. Those interested in CEO were encouraged by parole officers seek out services from the program. Under PFS, as eligible participants were about to be released they were assigned to the treatment group, and to individual parole officers who were instructed to make participation in CEO a condition of parole. Outreach staff from CEO met with PFS participants at their initial parole meetings to provide a bridge to CEO. Under the PFS pilots, CEO staff attempted to set up an appointment for recently released parolees to visit CEO.

**Job-readiness training.** After enrollment, participants were immediately placed in a five-day life skills education class. Taught by CEO instructors, these classes focused on resume writing, job-search skills, personal presentation, communication, appropriate workplace behavior, job interview etiquette, and techniques for answering questions about a criminal record.

**Transitional employment.** Upon completion of the life skills class, participants were placed into short-term transitional employment. To participate in the work crew, individuals had to call central phone number to indicate they were available for, and interested in, a work assignment. They were then assigned to work crews based on availability but are not guaranteed a spot on a work crew. Work crews performed maintenance, grounds-keeping, and custodial and repair work for city and state agencies and received a paycheck at the conclusion of every day they worked. The crews were supervised by a CEO staff member who ensured that tasks were completed and also identified inappropriate workplace behavior. CEO staff also identified individuals who might have required disciplinary action and also provided daily feedback on performance.

**Staff support while in transitional employment.** At the end of the first week, participants returned to the CEO office to work with staff who assisted the participants in accessing public benefits and community services (e.g., housing assistance, substance abuse treatment). This support was intended to help participants remain focused and motivated through their transitional employment period. After two weeks in transitional employment, CEO staff assessed participants for job readiness. Staff report that participants worked approximately 75 days across four months in subsidized transitional employment before being deemed job-ready.

**Job placement assistance.** Once determined to be job ready, participants worked with a job developer who helped them find full-time, unsubsidized employment. Job developers also contacted employers to identify potential job opportunities.

**Post-placement follow-up.** CEO staff provided participants with post-placement services up to one year after they attained unsubsidized employment. These included career planning, financial counseling, crisis management, and access to educational and vocational opportunities. CEO also offered monthly bonuses to participants (known as “rapid rewards”) for maintaining full-time employment and hosted retention events twice a year where staff and participants came together to discuss employment and other issues.

### **3.2.2 Changes in the CEO Program for the PFS Pilot**

In its PFS grant application, New York State described PFS as an opportunity to build on the MDRC evaluation of CEO’s programming that found the largest impacts among those who enrolled shortly after release from prison, and for those who were identified as being at high risk of recidivism by the Corrections Department. The state proposed to use PFS funding as an opportunity “to ensure that employment interventions are offered to the right people (those at high risk of reoffending) at the right time (soon after release from prison) to achieve the desired social outcomes (reducing recidivism, increasing attachment to the labor force, and containing costs in the public sector).” (State of New York 2013, 7)

To achieve the goal, CEO proposed and implemented a “client matching process” that identified and randomly assigned individuals who met specified criteria as they were being released on parole. Compared to its pre-PFS participants, this resulted in targeting a population that was at higher risk of return to prison, had been released from prison more recently, and was assigned to participate in the CEO program as a “special condition” of parole, rather than to voluntarily seek it out. Below we describe how this change affected CEO’s operations.

As discussed, prior to PFS, participants voluntarily decided to seek out CEO’s services. Parolees may have been told about CEO by their parole officer, but there was not a systematic process for identifying those at high risk and referring them to CEO (if assigned to the treatment group), and inducing them to engage with CEO, and to do so quickly. Changes to referral and outreach resulted in the following:

**Increased resources dedicated to outreach.** In NYC, CEO increased the number of outreach specialists from two to five. Outreach Specialists in NYC were assigned to regions around NYC and were responsible for following up with cases at specific bureaus that pertain to their assigned regions. Outreach began with the Outreach Specialist attending a “Matched Candidate Meeting” at the parole bureau, generally during a participants’ first meeting with their parole officer (PO) after release. At this meeting, the participant, the PO and an Outreach Specialist discussed the program, and the PO

explained the “special condition” which indicated that parolees were to go to CEO within two weeks of their release. The special condition was designed to drive participants to CEO, although finding the participant in violation of parole was at the discretion of the parole officer and generally depended on a number of other factors as well. Some parolees’ perception was that they were mandated to go to CEO because they were selected as part of a lottery.

During the meeting, the Outreach Specialist explained what the parolee would expect at CEO and scheduled a date for when the parolee was to arrive to CEO to start services. However, the meeting also served as an opportunity for Outreach Specialists to learn about potential barriers or issues the parolee would have to overcome with their release, including drug treatment, settling into housing and childcare, and any conflicts that would affect the parolee’s ability to show up to orientation. In Rochester, there was a similar increased focus on outreach and formalization of the referral process by the parole office. There was only one Outreach Specialist due to the size of the site, but he also attended matched candidate meetings, frequently being notified as a participant was showing up at the parole office and meeting with them the same day. The Outreach Specialist then continued contact with the participant to encourage them to attend orientation.

**Enrollment of a more challenging group of participants.** Participants identified through PFS were those classified in the two highest risk categories on a four-point scale indicating risk of reoffending, and some were randomly selected to be in the treatment group. Outside of the treatment group members in the PFS pilot, any recent released parolee could seek out CEO. PFS participants were also pushed to engage sooner after release (partly due to the state’s articulated goal of offering services soon after release, and partly in an effort to meet project enrollment and outcomes under the grant’s time frame). However, according to CEO staff, some PFS treatment group members were still transitioning to life outside of prison and were not ready or motivated to commit to the time required for participation in CEO. Finally, CEO participation being identified as a special condition of parole gave parolees the sense that the program was mandatory. This resulted in participation from individuals who might not otherwise be motivated, but also, among those who might have attended anyway, there might have been a difference in motivation based on the perception that participation was required. Focus group participants noted that their initial perception of CEO was negative because they thought that it was a “program” (similar to mandated drug treatment or anger management programs), but many felt more positively about it once they began to view it as an employment agency.

**Increased communication with the parole officers.** The addition of the “matched candidate meetings” at parole offices meant that CEO staff had a more structured relationship with individual parole officers. CEO staff also noted that attention of PFS outcomes from the state may have reached individual parole officers, and that they were more invested in parolees’ participating in CEO. If a randomly assigned participant stopped showing up for services, for example, there was now an avenue for CEO to follow up directly with the parole office. According to CEO staff, once they got to know parole officers, they were able to do this whether the individual they were concerned about was a PFS participant or not.

**Increased focus on tracking participants.** CEO has traditionally had strong data systems and was selected for participation in PFS in part because of their familiarity and comfort with data. Participation in the PFS pilot bolstered this focus. There were regular meetings between CEO and DOCCS dedicated to reviewing lists of PFS participants and tracking whether they were attending CEO and progressing through the project. Within CEO, staff knew if participants were assigned

through PFS. In addition to outreach staff making a concerted effort to engage PFS participants and increased efforts to re-engage participants who dropped out, employment retention staff paid special attention to participants around the “PFS quarter” (the quarter in which the employment outcome was measured). There was also a sense among staff and participants that PFS participants got priority in getting a work crew assignment because the participation in transitional employment outcome was tracked so carefully. Work crew supervisors noted that PFS led to a somewhat different approach to discipline in that there was greater motivation to keep participants on work crews even if they violated program policies.

**Changes in the outcomes of focus.** The outcome targets related to recidivism in the PFS pilot were largely aligned with the outcomes that CEO had previously tracked. The employment metric under the PFS pilot, however, was seen as meaningfully different from prior practice. Generally, CEO was focused on employment outcomes for 365 days after a participant found a permanent job (as opposed to placement on a transitional work crew). This allowed the organization to work with participants for varying lengths of time to get them “job ready,” and also allowed for the fact that participants may enter the program at various points after release. Staff reported that the focus on employment in the fourth quarter after *release* meant they had to move participants through the CEO program more quickly. Some staff felt this shifted the focus on initial job placement rather than sustained employment.

### 3.3 Summary

Both PFS pilots engaged service providers that worked with individuals involved in the criminal justice system and focused on employment and reducing recidivism. Roca had an intensive four-year model that aims to “break the cycle” of poverty and incarceration by making use of transformational relationships and cognitive behavioral approaches. CEO sought to use transitional jobs to help recently incarcerated individuals build job skills and an employment history. Both service providers were well-established and had been operating their programs for many years and had experience using data to track outcomes. CEO had been the subject of a previous RCT study that found positive impacts for higher-risk participants.

Both organizations operated the same service models under PFS pilot as they had before, but there were some differences under the pilots. Roca expanded their service area and opened several new sites, while CEO attempted to serve high risk individuals who began services shortly after release. For both service providers, participation in PFS led to changes in the referral and recruitment process.

## **4. Random Assignment and Enrollment Challenges in the PFS Pilots**

Identifying eligible participants, randomly selecting them for the treatment or control group, and enrolling treatment group members in program services were critical tasks under the two pilots. Also critical was reaching sufficient enrollment levels during a specific time period under the DOL grants, so that services could be delivered, enrollment targets could be attained, and employment and recidivism outcomes detected, which, if successful, would trigger payment from DOL. This chapter describes how random assignment and referrals were conducted in the two pilots, the challenges that occurred in meeting enrollment targets, and adjustment efforts undertaken.

### **4.1 Implementation of Random Assignment in the Massachusetts Pilot**

Phase I of the Massachusetts pilot was aligned with the four-year period of performance for the DOL grant and required the pilot to complete start-up and implementation activities within the first 36 months. The project had a goal of enrolling 535 individuals in the Roca program over a two-year random assignment period. Recognizing that not all individuals assigned to the treatment group would enroll in services, the Massachusetts partners estimated that the target would be met by randomly assigning 753 individuals to the treatment group over the first two years of the project (a 71 percent enrollment rate).<sup>10</sup> The pilot began random assignment and enrollment in Roca services in January 2014, three months after grant award. Enrollment for Phase I continued through December 2015, and outcomes for these enrollees were tracked through September 2016

#### **4.1.1 Overview of the Random Assignment Procedures**

For the Massachusetts PFS pilot, the steps involved identifying eligible participants and randomly assigning them to the treatment and control groups (shown in Exhibit 4-2) and providing information to Roca on potential treatment group enrollees. This included the following:

- Referrals for the PFS pilot were sent monthly from multiple agencies to the state's Office of Administration and Finance. The referring agencies included the Department of Youth Services, Office of Probation, Massachusetts Department of Corrections, the Massachusetts Parole Board, and corrections agencies in Suffolk County, Essex County, Hampden County, and Middlesex County. The referral partners electronically provided lists of individuals who appeared to meet the intervention's eligibility criteria. The criteria were that individuals must be male and age 17 to 24; be medium or high-risk based on a risk assessment; must be aging out of Department of Youth Services or on adult probation; and must be living in the Boston, Chelsea, or Springfield area.

---

<sup>10</sup> As discussed in Chapter 2, the PFS pilot in Massachusetts initially contracted with Sibalytics, to implement and oversee the random assignment process but in late 2014, replaced the company with the Urban Institute. This change did not affect the random assignment procedures that had been developed for the random assignment and data collection and analysis.



---

## RANDOM ASSIGNMENT AND ENROLLMENT CHALLENGES

---

- The Office of Administration and Finance sent the referral lists to Urban Institute. On a monthly basis, Urban Institute matched this list against records maintained by the Department of Criminal Justice Information Services to determine if any candidates were ineligible for the PFS pilot based on other criteria. i.e., that individuals had a prior enrollment in the PFS pilot, a prior enrollment in Roca, an open felony charge, or a conviction/adjudication due to a sexually violent offense are not eligible for the PFS pilot. All remaining individuals were then randomly assigned to the treatment or control groups by the Urban Institute, which in turn provided identifying and contact information for individuals assigned to the treatment group to Roca. The list was sent monthly to Roca via a secure electronic file.
- Upon receiving the list of individuals assigned to the treatment group, Roca assigned each treatment group member to a youth worker who attempted to contact the individual. If contact was established, Roca staff had 60 days to re-assess the individual's eligibility for services. Individuals who had new criminal charges since the time of random assignment or those who did not meet Roca's own eligibility criteria of a background indicative of future involvement in the criminal justice system (i.e., no high school diploma, limited or no work history) were not eligible.

### 4.1.2 Challenges in Identification and Referral of Treatment Group Members

The Massachusetts PFS pilot under Phase I experienced a number of challenges in identifying eligible individuals, and recruiting and enrolling them in a timely manner, as a result of the procedures to implement the RCT, which ultimately affected its ability to achieve the target enrollment level. These challenges were related to several factors:

**Cumbersome data transfer procedures** from the criminal justice agencies and courts slowed referrals (e.g., several different trial courts had to manually enter certain participant data, and some agencies provided incomplete and/or out-of-date data).

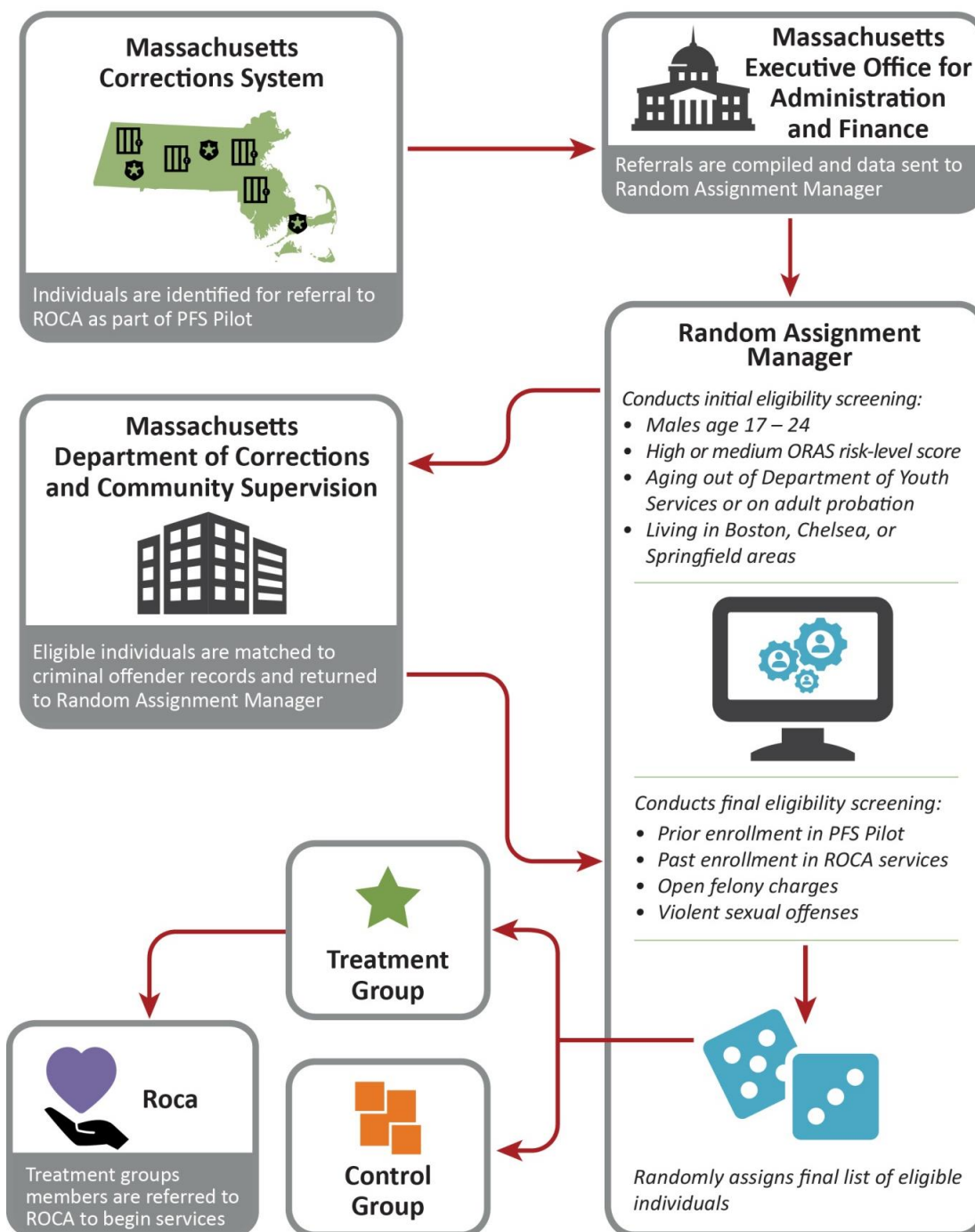
**Poor quality of participant contact information**, which hindered Roca's ability to find individuals and enroll them in the program. Roca staff estimated that one-third of the individuals who were referred to, but not enrolled in, Roca could not be located with the contact information given to them by the referring agencies. Roca staff reported that early in the project, many of the referrals they received were "stock" referrals of young men who had been on probation for a long time, and the contact information associated with these men was often outdated.

**Individuals randomly assigned who either were not eligible for the Roca program or were not able or willing to participate in it.** Some of the individuals referred to Roca were not at the high-risk level required by the program and therefore were ineligible. By Roca's criteria, to be eligible for services, individuals must be both willing to make a commitment to the program and demonstrate a need for the program services. Individuals are only eligible for Roca if they are not able to hold a job, successfully engage in another program, or show up and engage in Roca programming from the start. However, due to the procedures for random assignment, a participant's willingness to participate in Roca services could not be determined until after random assignment was conducted. In addition, some participants were determined to be ineligible because they did not reside in one of the three service areas, were still in prison or had been re-arrested, or they had an open charge that was not known to Roca until after random assignment. Further, Roca also did not serve individuals with severe mental health, developmental delays, or substance use disorder. PFS partners reported that it

## RANDOM ASSIGNMENT AND ENROLLMENT CHALLENGES

was not possible to screen out individuals with these conditions because those data are not collected in any existing state or county databases.

**Exhibit 4-1: Random Assignment Procedures in the Massachusetts PFS Pilot**



Low enrollment also resulted from an overestimate of the number of potentially eligible individuals. The partners in the Massachusetts pilot reported that they did not know what was behind the projection error but speculated it could be due to reduced criminal activity or changes in sentencing practices.

### 4.1.3 Efforts to Address Low Enrollment in the Massachusetts Pilot

Partners notified investors of implementation challenges through the formal channels as per the contract and reporting structure, and investors reported that they were aware of most problems, particularly those related to referrals, from the start. During the initial months of the pilot when referrals were lower than anticipated, Operations Committee members reported frequent meetings that required a significant time commitment. Communication during this time sometimes happened daily. The state of Massachusetts, Third Sector, and Roca reported that the management structure worked well in the pilot and that all parties worked collaboratively to address the challenges. Together, pilot partners developed a series of adjustments to improve referrals to and enrollment in the Roca program. These included the following:

- To improve the timeliness of referrals, the Massachusetts pilot worked with Office of the Commissioner of Probation to provide its data through MassCourts, which reduced the time between referral data transmissions from 90 to 60 days.
- To increase the number of eligible individuals, partners expanded the age range of individuals served by Roca from 22 to 24 years old at the top end and broadened the eligibility requirement to allow medium-risk individuals to enroll.
- Partners added six new state referral agencies: the Massachusetts Department of Corrections (DOC), the Massachusetts Parole Board (MPB), and sheriff's departments in Suffolk County, Essex County, Hampden County, and Middlesex County.
- To obtain participants that more readily met the eligibility requirements, pilot partners tried to determine how to screen out individuals who were gainfully employed or enrolled in a formal education program but were not successful.

Roca and the other pilot partners developed several strategies for addressing the problem of poor contact information for referrals.

- Third Sector and Roca worked proactively with staff in Probation and DYS to improve efforts to verify addresses and other information before submitting referrals to Roca. Roca staff reported also participating in weekly meetings with Probation to discuss referrals.
- Roca developed agreements with DOC and the four sheriff's departments so these agencies would notify Roca about referrals prior to their release, allowing Roca to begin the outreach process before these individuals were released from custody.
- PFS partners extended Roca's outreach period from 60 to 90 days to give Roca staff more time to locate referrals. Finally, Roca adopted a new internal policy requiring youth workers to conduct outreach with a supervisor at least twice before designating a referral as non-conforming. Roca also instituted weekly referral meetings internally to maintain momentum in locating referrals.

Roca also started its own direct recruitment rather than relying on referrals from other agencies. The pilot contract among the state of Massachusetts, Roca, and Third Sector permitted Roca to "self-

## RANDOM ASSIGNMENT AND ENROLLMENT CHALLENGES

---

recruit” participants if referrals are insufficient. Roca began recruiting participants using the same approaches used prior to launching PFS. Because they were not referred from the criminal justice system, individuals recruited in this way are not counted in the impact evaluation assessing recidivism and employment and earnings, they are counted toward the job-readiness outcome (see Chapter 5).

In December 2014, several of these adjustments were formally documented in a modification of the project’s contract between Massachusetts, Roca, and Third Sector (“Amended and Restated Pay for Success Contract”). The contract also extended the enrollment period for the pilot by two quarters and reduced the projected number of referrals and enrollments. (The change in the duration of the enrollment period affected timing of success payments made by Massachusetts in Phase II of the pilot, but not any payment by DOL as part of Phase I.)

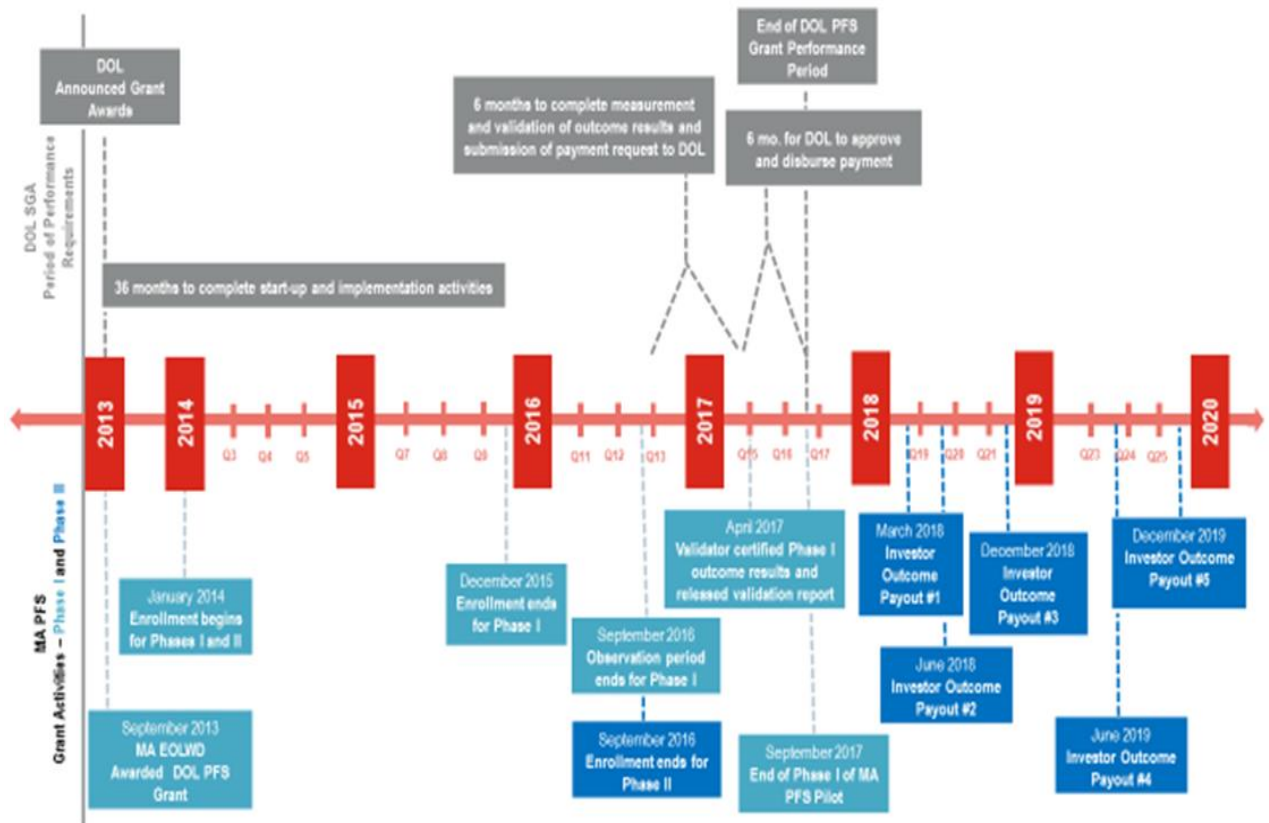
Despite the strategies employed above, the pilot continued to face shortfalls in referrals and enrollment. In 2016, when the Massachusetts pilot failed to refer a sufficient number of eligible individuals to Roca for three consecutive quarters, an early termination clause in the agreement between Third Sector and the investors was triggered, allowing the investors the right to call their loans into default. During this period, all PFS partners, including the investors, met weekly by phone, and at least twice in person.

After approximately a year of frequent meetings investors decided to waive the clause and continue funding the pilot. The partners agreed to reduce the number of referrals required and extend the enrollment period for an additional six quarters. The Arnold Foundation agreed to fund the extended enrollment period, including the costs of providing services to participants, evaluating results, and project management through a grant to Roca. These changes were documented in a formal contract modification (“Second Amended and Restated Pay for Success Contract”) signed in November 2016. Investors reported that two factors convinced them to continue funding the PFS pilot: the commitment demonstrated by Massachusetts to make the project work and investors’ desire to support Roca by funding services for participants and the program evaluation.

Based on data provided to DOL, during Phase I of the pilot, 991 individuals had been referred to the treatment group, with 242 of those individuals enrolling in Roca, (a 24 percent enrollment rate), and well short of their target of enrolling 535 Roca enrollees.

Phase II enrollments began in January 2016, with plans for enrollment to end in September of that year, and for the observation period to end in June 2019. The project partners for the Massachusetts pilot did not differentiate between the two phases of the pilot, but rather considered the project in its entirety for purposes of meeting various targets and possible payouts. A timeline (in Exhibit 4.2) for the entire pilot identifies the milestone under both phases.

Exhibit 4-2: Phase I and Phase II Milestones in Massachusetts PFS Pilot



**4.2 Implementation of Random Assignment in the New York PFS Pilot**

The DOL PFS pilot in New York had the same four-year performance period as did Massachusetts, from the date of the grant award in September 2013, with 36 months for start-up and implementation activities, including random assignment, recruitment, enrollment and delivery of the intervention, and six months for measurement, documentation, and validation of outcome results. (DOL had a final six months of the grant to review validated results and potentially to approve and disburse payment.)

In the New York pilot, the Research Office of the state’s Corrections department was responsible for identifying prospective participants, conducting random assignment, and providing information on treatment group members that would be passed along to parole officers. Random assignment started in December 2013, and continued until September 2015 under Phase I, with an enrollment target of 1,000.

**4.2.1 Random Assignment Procedures in the NY Pilot**

Procedures for random assignment (summarized in Exhibit 4-4 below) involved the following:

- DOCCS maintained an updated file of prisoners nearing release. Two to four weeks before a prisoner’s release date, DOCCS examined his eligibility for the CEO program based on two

## RANDOM ASSIGNMENT AND ENROLLMENT CHALLENGES

---

factors: the location of the individual's release from incarceration and his supervision level.

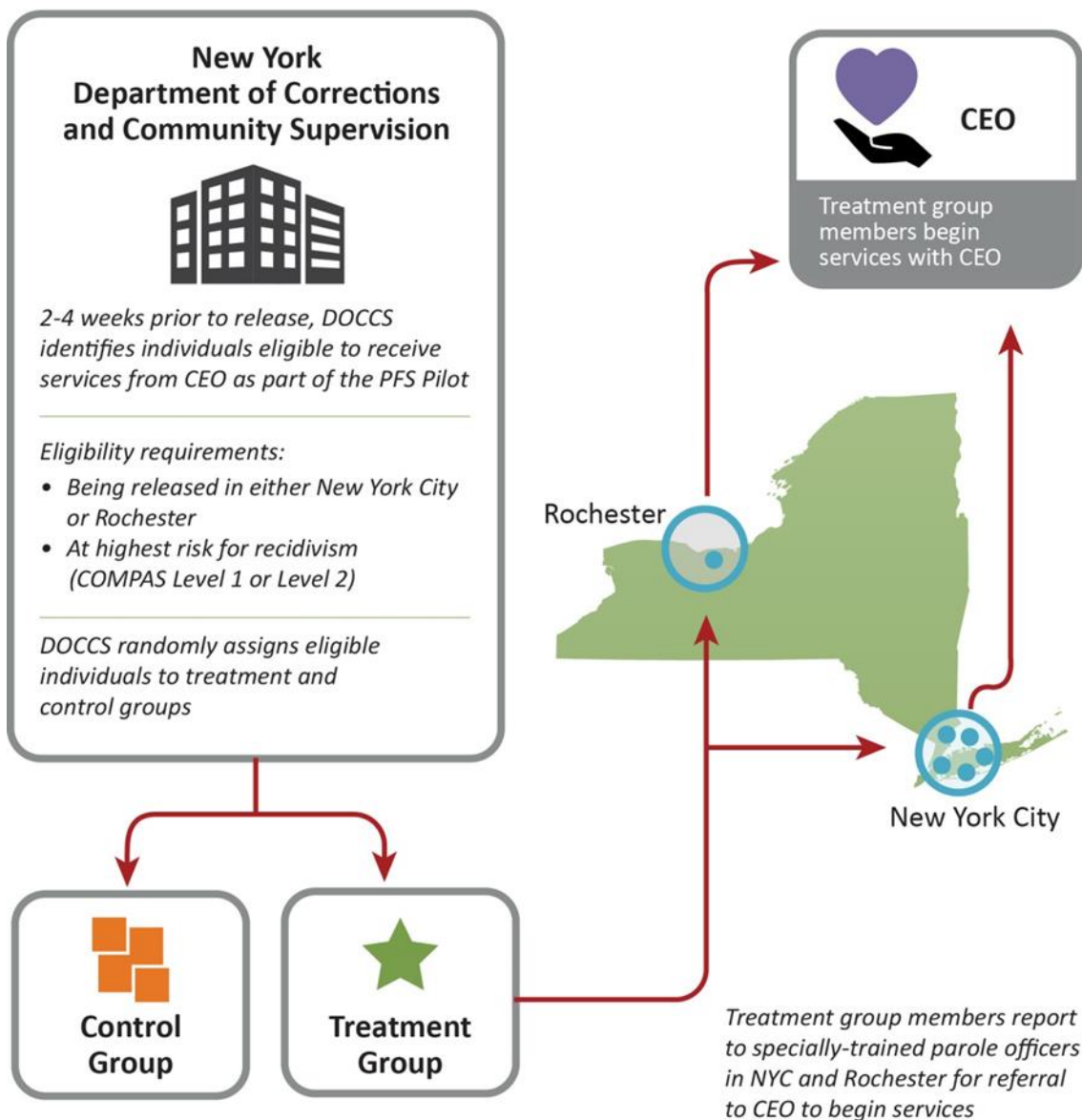
Eligible individuals were those released to community supervision at the locations participating in the pilot: one of the five New York City parole bureaus, the Queensboro Correctional Facility, and the Rochester Metro parole bureau. Additionally, potential participants had to be assessed as being at the highest risk for recidivism under the state's assessment system and had to agree to participate in the study and the services as part of the terms of their parole (individuals who do not consent are not referred to CEO or included in the study).

- Those meeting the eligibility criteria were randomly assigned by research staff at DOCCS. DOCCS emailed the identity of treatment group members to a specific parole officer who was responsible for all individuals in the PFS pilot and to the service provider, CEO. Additionally, treatment group members were flagged in the management information system used by parole officers and maintained by DOCCS to ensure they were referred and served appropriately.
- As part of their parole requirement, treatment group members had to contact their designated PFS parole officer within 24 hours after release and schedule an in-person meeting with the parole officer within a week. This meeting also was attended by a member of the CEO staff with the goal of enrolling the individual in the CEO program. PFS parole officers received special training on the benefits of the CEO program and how to make successful referrals. The PFS parole officers also were provided scripts to help them encourage parolees to enroll at CEO. To help minimize the possibility of control group members accessing CEO services on their own, control group members are assigned to parole officers who have not received the training on CEO and who had a limited history of referring parolees to CEO in the past.

The steps involved in implementing the RCT required significant procedural changes in terms of how individuals were referred to PFS. As discussed in Chapter 3, eligible PFS participants were identified by DOCCS were directed to parole officers selected to participate in the pilot and instructed that participation in CEO was a special condition of their parole. Outreach staff from CEO joined the initial parole meetings with PFS participants to provide a bridge to CEO services. This recruitment strategy was new to both DOCCS and CEO. Prior to the PFS pilot, parole officers used discretion in referring parolees to CEO services and did not directly coordinate meetings with CEO staff. This change in the referral process formalized the relationship between the parole bureaus and CEO and led to reoccurring joint staff meetings during the PFS pilot to review the list of treatment group members and schedule CEO intake appointments.

The New York pilot met its target of 1,000 for the two-year enrollment period for Phase I, with 1,502 individuals randomly assigned to the treatment group, of who 1,068 enrolled in CEO services (a 71 percent enrollment rate).

Exhibit 4-4: Random Assignment Procedures for the New York PFS Pilot



**4.2.2 Challenges in Random Assignment and Enrollment**

While the New York pilot met its enrollment goals, it did experience challenges in reaching those goals. The key issues encountered, and how the pilot partners addressed them, are discussed below.

Early supervision status reports released by DOCCS Research showed that the actual rate of eligible releases did not match the sample size estimates in New York City or Rochester. Members of the governance committees agreed that the projected number of releases would be insufficient for achieving the target sample size of 1,000 participants in Phase I.

## RANDOM ASSIGNMENT AND ENROLLMENT CHALLENGES

---

To detect a difference in outcomes between the treatment and control group, as discussed in Chapter 2, the evaluation design required the proportion of treatment group members receiving CEO services to be at least 30 percentage points higher than the proportion of control group members receiving services similar to those received by the treatment group. The final Phase I enrollment numbers showed that a total of 1,502 individuals were identified as eligible, but 1,222 were referred to CEO and 1,068 actually attended an orientation at CEO. In addition, of the 855 individuals in the control group, 567 were referred to CEO, and 431 attended an orientation. Although the treatment and control groups' contrast met the criteria, the number of treatment group members enrolled in CEO was lower than expected level, and a greater number than anticipated of control group members had enrolled in the CEO program on their own.

The New York governance committees monitored the pilot and identified various intervention and enrollment challenges by reviewing periodic reports that were provided by DOCCS. These reports raised questions about whether it was possible to reach the target sample size and show a treatment contrast during Phase I of the pilot. In response, governance committees continued to meet to review operations, discuss issues, and propose solutions for the challenges encountered by the PFS pilot. Exhibit 4-5 shows that partnership management implemented changes to the evaluation and intervention during Phase I of the pilot to increase the total sample size and treatment contrast. These changes included:

- The addition of two parole bureaus in New York City in April 2014 and an adjustment to the random assignment ratio in Rochester to increase the number of parolees assigned to the treatment group and decrease the number assignment to the control group in June 2014. In December 2014, another three parole bureaus were added, bringing the total number of parole bureaus participating in New York City to nine.
- To increase treatment group enrollment into CEO services, in October 2014 the working groups approved a process that allowed CEO staff to bring MetroCards (New York City public transportation passes) to initial meetings with parolees assigned to the treatment group. The MetroCards served as an incentive for parolees to enroll in CEO.
- In September 2015, the governance committees approved a change to permit parole officers the flexibility of meeting with parolees directly at CEO facility instead of at the parole office. The intent of this change was to increase parolee enrollment in CEO and retain enrollees in the pilot.
- To maximize the number of treatment group members enrolled during Phase I, a one-month extension of the observation period was provided by the governance committee. The extension also included revising the measurement start date and validation period start date.

Although the governance process operated as planned, some partners expressed concerns with their limited ability to implement changes to the pilot beyond what was specified in the state agreement with the partners, and some partners reported that New York workforce agency was unwilling to renegotiate certain aspects of the agreement and only allowed minor modifications. The state agency shared that the agreements were not intended to be flexible, and Harvard's GPL also emphasized the importance of limiting modifications based on preliminary data, as that would jeopardize the evaluation. Also, proposed changes to the random assignment or evaluation design required Executive Steering Committee approval as well as an amendment to the partnership agreement. The independent validator was also required to review and approve any change related to the evaluation

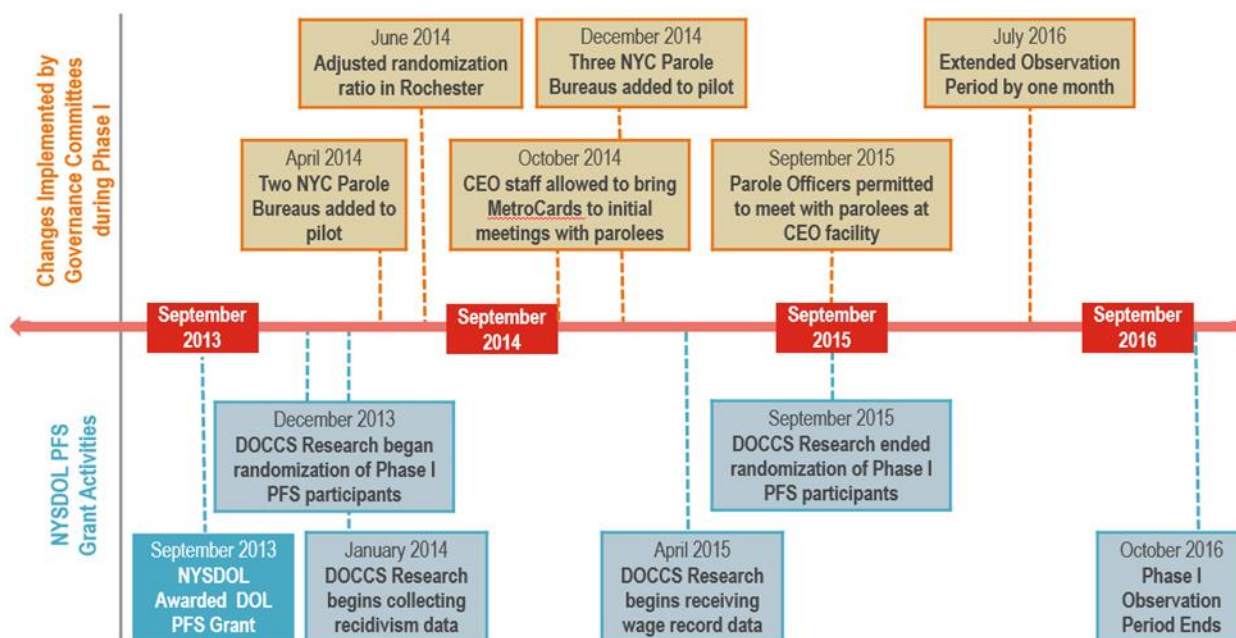


## RANDOM ASSIGNMENT AND ENROLLMENT CHALLENGES

design and measurement of outcomes. Proposed changes to randomization, or the timing and content of reports required Management Committee approval, and proposed changes to parole staff or the workload of DOCCS needed approval by the working groups and relevant organization division head.

While these requirements may have limited some actions that the partners could take, the state agency reported that the PFS pilot was one of the most engaged stakeholder partnerships in the recent history of the Department and attributed the strong engagement to the governance structure. The New York pilot partners were in regular communication and understood that the pilot required regular participation from all stakeholders in order to effectively implement the evaluation.

### Exhibit 4-3: NYS Governance Committees Approved Operational Changes to the PFS Pilot during Phase I



### 4.3 Summary

Both Massachusetts and New York faced and addressed a number of unexpected challenges in operating the PFS pilots, particularly in random assignment and referral procedures. The management and partnership structure that the pilots put into place assisted each to meet the challenges. The Massachusetts PFS pilot experienced a number of issues in receiving timely and viable referrals and made a number of adjustments to increase referrals and enrollments as well as changing the evaluator, but still fell short of target enrollments. The New York PFS pilot experienced a lower than expected number of eligible releases (their pool of possible participants) and thus enrollments in the treatment group were not as high as estimated initially, although their sample target was met after some adjustments. Some partners found that the PFS contract agreements, while providing directives on operations, limited flexibility and timely decision making. Overall, the PFS partners in both states reported being highly engaged and this engagement helped to address the issues that arose.

## 5. Evaluation Findings and Payment Results

A central tenet of the PFS approach is that investors who help fund new service interventions will earn a return on their investment if the new social program or policy is more successful than current approaches. Investors will share in the resulting savings to the public if the key outcomes of the intervention are better and/or less expensive to achieve than “business as usual.”

A major challenge for any PFS project is the accurate determination of intervention impacts, or whether it has succeeded in meeting its goals in improving social outcomes. As discussed in Chapter 1, previous PFS projects have been subject to criticism because of a lack of confidence that the evaluation methodology used provided an accurate assessment of the project’s impacts. In particular, the failure to ensure a valid comparison between the new intervention’s results and what would have happened in its absence is a commonly cited limitation of earlier PFS implementations.

DOL’s SGA took steps to address these concerns by requiring: (1) that projects use a “credible quasi-experimental” or experimental evaluation design to make this assessment; and (2) that grantees contract with an independent validator to assess the integrity and validity of the evaluation and its findings.

The evaluators implemented the study as planned and agreed to by PFS partners in the grant applications. Moreover, the study validators independently verified the results. In this chapter, we discuss the measures used to determine whether the intervention produced the target impacts; the evaluation findings and validation approaches and results; and grantee decisions about requesting DOL payments for investors.

### Evaluation Findings and Payment Results: Key Points

- *Both pilots assessed impacts on recidivism and employment, and measured outcomes on engagement in services.*
- *Neither pilot found statistically significant positive impacts as a result of the intervention for the federally funded first phase of the PFS projects.*
- *Both pilots engaged validators, and the validators certified the evaluation’s findings, but the two pilots interpreted the validator’s roles differently.*
- *While DOL took important steps to ensure stated evaluation results were backed by credible methodology, carrying out the evaluation and validation functions were the primary responsibility of the grantee, and as a result, the grantees carried out these responsibilities differently from one another.*

### 5.1 Measures of the Intervention’s Effects

Both PFS pilots designed and implemented RCTs to measure intervention impacts on two key outcomes for the target population: recidivism and employment. The pilots also measured project performance in a third area: the level of engagement of the treatment group in the service

## EVALUATION FINDINGS AND PAYMENT RESULTS

---

intervention. As planned, DOL was to use these findings as indications of potential government savings and grounds for making payments based on the findings.<sup>11</sup>

Both PFS pilots specified the outcomes of interest as follows:

**Recidivism.** To measure the effectiveness of the service interventions in reducing recidivism, both pilots measured the impacts by comparing the treatment and control groups' average days they were incarcerated over a four-year period. This measure was identified as "bed-days avoided," meaning that if the intervention kept parolees from re-offending it might reduce future prison populations and generate savings in incarceration costs.

To measure the effects of the service intervention on reducing recidivism, the Massachusetts and New York PFS pilots measured the impact of the service intervention on the days of incarceration resulting from new offenses committed during the first five years post-release. The estimates of days of incarceration were based on length of the sentences issued for treatment and control group members during the study period, scaled up to five years, with imputed sentences for any open arraignments that had occurred but not yet been fully adjudicated by the end of the study period. The impact, or the difference between the treatment and control group on this measure, provided a measure of bed days avoided or saved due to the intervention.<sup>12</sup>

**Employment.** To measure the effectiveness of the service interventions on employment, both pilots measured the impact of their program on the employment rate of treatment group members compared to that of control group members. The increased level of employment and related earnings was hypothesized to provide budgetary savings through increased taxes and reduced public assistance receipt.

To examine whether the service intervention increased employment levels, the Massachusetts PFS pilot used a measure of post-release employment and earnings. Specifically, the measure was the percentage of treatment and control group members who were employed with earnings of more than \$1,000 each quarter during the study follow-up period. Pilot partners reported that the employment threshold measure of \$1,000 earned in a quarter was equivalent to working half-time in a relatively low-paying position and was selected to focus on achieving a "substantial" level of employment. Again, the impact, or the difference, between the treatment and control group on this outcome, provided a measure of the increase in employment due to the service intervention.

To examine whether the service intervention increased employment levels, the New York pilot used a measure of post-release employment and earnings. Specifically, the pilot measured the impact, or the difference, between treatment and control group members, on employment rates in the fourth quarter

---

<sup>11</sup> This study's interim report includes more details about how the evaluators measured potential savings. See Chapter 7, Abt 2016.

<sup>12</sup> For the recidivism measure, the evaluation plan anticipated the challenge of missing data. It is important to note that when a study is missing outcome data for some subjects, researchers may choose to replace the unobserved data with data that have been imputed in some way, rather than exclude subjects with missing outcome data from the analytic sample. This practice is common and there are accepted ways to impute outcome data (e.g., regression imputation, non-response weights, maximum likelihood estimation, and dummy variable imputation).

after release from incarceration. This measure was selected because prior research on CEO's program shows this time period is typically when the highest nonsubsidized earnings are reported.

**Engagement in services.** Each pilot also established a measure of the extent to which enrollees in the treatment group participated in the services offered by the service provider (with no comparison to a control group). The New York PFS pilot used this measure to calculate budget savings, and in Massachusetts, this measure was used in determining payouts to investors using state funds. Neither grantee reported this outcome for Phase I, nor is it discussed below.

### 5.2 Evaluation Results and Validation

DOL's SGA specified that each pilot include an evaluation and an independent validator of evaluation results. The two grantees carried out these tasks somewhat differently. Massachusetts engaged an external evaluator and separate independent validator. New York conducted the evaluation internally through the DOCCS research office and engaged an independent validator to certify the results. In this section we describe Phase I evaluation results as well as the validator's approach and findings.

#### 5.2.1 Evaluation Results and Validator Findings in Massachusetts

The Urban Institute conducted the impact analysis in March 2017 (three and a half years after enrollment began), examining outcomes for youth randomly assigned from the second quarter ending March 2014 through the ninth quarter ending December 2015 of the study (no random assignment occurred in the first quarter ending December 2013). During this period, 991 individuals were randomly assigned to the evaluation: 748 to the treatment group and 243 to the control group.

Among those assigned to the treatment group, 32 percent (242 individuals) enrolled in Roca and among control group members, 5 percent (13 individuals) enrolled in the program.<sup>13</sup> Data were collected through Quarter 11, June 2016, for the employment outcome and Quarter 12, September 2016 for the recidivism outcome (e.g., data covered between about 6 months and 2 ¼ years for employment, and 9 months and 2 ½ years for recidivism). Young men randomized in the second quarter ending March 2014 were observed for the greatest number of quarters (nine for recidivism and ten for employment) and those randomized in the ninth quarter ending December 2015 were observed for the smallest number of quarters (two for recidivism and three for employment). On average, participants were observed for 5.8 quarters for the employment outcome and 7.3 quarters for the recidivism outcome.

Urban provided two estimates for each outcome – Intent-to-Treat (ITT) estimates and Treatment-on-the-Treated (TOT) estimates. The ITT analysis compared the outcomes of all individuals assigned to the treatment group to the outcomes of all individuals assigned to the control group, thus estimating the effect of *access* to pilot services. The TOT analysis measured the impact of intervention services

---

<sup>13</sup> Note that it was possible for control group individuals to self-refer to Roca and, if otherwise eligible, they were not turned away.

## EVALUATION FINDINGS AND PAYMENT RESULTS

for individuals who received grant-funded services, thus estimating the effect of *receipt* of pilot services.<sup>14</sup>

As illustrated in Exhibit 5-2, neither estimate showed a statistically significant difference at the 95 percent confidence level or better between the treatment and the control group – an impact – on recidivism, as measured by average bed days, nor on employment, as measured by the average number of quarters employed, defined as reported income of at least \$1,000 in a given quarter.<sup>15</sup> To achieve an adequate level of cost savings, the Massachusetts pilot sought a reduction of at least 100 bed-days. The level of employment impacts was not specified, but they were required to be statistically significant.

### **Exhibit 5-2: Summary of Phase I Recidivism and Employment Outcomes for Massachusetts PFS Pilot**

Outcome	ITT Estimates		TOT (IV) Estimates	
	Impact	Standard Error	Impact	Standard Error
<b><u>Recidivism Outcome</u></b>				
Average Bed Days	13.6	17.6	50.0	64.0
<b><u>Employment Outcome</u></b>				
Average Quarters Employed	-0.04	0.16	-0.14	0.54

SOURCE: Massachusetts Social Innovation Financing Youth Recidivism Pay-for Success Initiative Independent Validator, Interim Outcome Estimation Validation Quarter 14 Report (Quarter Ending March 2017)

Acting as the validator, PCG reviewed Urban Institute’s evaluation findings memo and the files they used to produce the estimates. In comparison to the independent validator in New York (see section below), PCG took a narrower role of reviewing the data files and programming to ensure that the analysis files were programmed correctly and that the evaluation’s results were reported correctly.

They did not explore broader topics that could have affected confidence in the evaluation findings such as whether the integrity of random assignment was maintained and whether data were measured and collected as planned. Specifically, two issues affect the interpretation of the evaluation results. First, as discussed in Chapter 4, the size of treatment group enrollment was lower than originally expected due to challenges with referrals and eligibility for the program. Second, the length of the observation period for Phase I may not have been sufficient to capture the program’s impacts. For example, Roca is a four-year intervention. During the first two years, participants receive intensive services and in the last two years, they receive less intensive, follow-up services. Given the schedule of sample entry, no individuals in the DOL-related evaluation sample had time to be observed for impacts at four years after random assignment and few had the time needed to be observed for

<sup>14</sup> Note that the TOT estimate relies on predictions about who among the control group *would have* engaged in ROCA services had they been in the treatment group. TOT estimates were not based on an experimental framework, but instead use the instrumental variable (IV) approach to predict program participation for the control group. An instrumental variable predicts participation without predicting outcomes, allowing for an unbiased prediction of who in the control group would have participated in the intervention.

<sup>15</sup> As described in the memo included as part of the validator’s report, the evaluator averaged employment data across small groups of individuals rather than for individuals in the sample. Individuals were grouped into 152 groups and regressions were run at the group level to examine the impact on employment.

## EVALUATION FINDINGS AND PAYMENT RESULTS

impacts even at two years after random assignment. Consequently, even if all treatment referrals had been fully engaged in Roca for the lifetime of the intervention, most would not have had time to complete even the first two years of the intervention before evaluation findings for Phase I were reported.

PCG reported that its involvement in the project was limited. The validator participated in quarterly meetings of the Oversight Committee and validated Urban’s work on a monthly basis. PCG reported that it was not notified about issues with the project (e.g., challenges with referrals, low enrollment) until it reviewed Urban’s data or participated in quarterly project meetings.

### 5.2.2 Evaluation Results and Validator Findings in New York

The PFS impact study measured differences in employment and recidivism outcomes for a sample of 1,502 individuals randomly assigned to a treatment group referred to CEO services and for a control group of 855 that were not directly referred to CEO services, although they were able to seek out services at CEO or other organizations on their own.<sup>16</sup> As discussed in Chapter 2, New York developed a back-stop evaluation strategy to be used if the proportion of treatment group members receiving CEO services was less than 30 percentage points higher than the proportion of control group members receiving CEO services. This back-stop methodology was not invoked as the difference in the proportion enrolled between the two groups was 43 percentage points.

As Exhibit 5-3 shows, the final outcomes of Phase I do not provide evidence that the New York pilot met its target employment or recidivism impacts. The evaluation concluded that the project did not achieve its stated goals of an employment rate five percentage points greater in the 4th quarter following random assignment or an eight percent reduction in average bed-days for the treatment group relative to the control group.

**Exhibit 5-3: Summary of Phase I Employment and Recidivism Outcomes for New York PFS Pilot**

Outcome	ITT Estimates			TOT Estimate
	Treatment Group	Control Group	Impact (Difference)	Impact (Difference)
<b><i>Employment Outcome</i></b>				
Employed in 4 <sup>th</sup> Quarter	17.6%	16.8%	0.8%	2.0%
<b><i>Recidivism Outcome</i></b>				
Average Bed Days	232.5	229.5	3.0	7.0

SOURCE: New York State PFS Project: “Employment to Break the Cycle of Re-Incarceration” Phase I Validator Report

Although the validator in New York, Mike Puma Associates, took on a broader role than the validator in Massachusetts, the firm did not receive access to the data or report on the analysis files. Per the firm’s contract with the state, Mike Puma Associates produced a 20-page report summarizing its role, activities, and findings as the validator. The report included an overview of the project, the role of the independent validator, a review of the project as planned, a review of the project as implemented, and the determination of outcomes. The role included communications over the course of Phase I, first in establishing the validation methodology and determination process, conducting an initial assessment

<sup>16</sup> Note that it was possible for control group individuals to self-refer to CEO and, if otherwise eligible, were not turned away.

of planned project procedures, monitoring implementation of the project and random assignment, and validating the calculation of the outcomes and impacts.

Mike Puma Associates conducted its work through site visits and interviews with project staff, as well as a review of project documents. The validator's report describes initial areas of concern around the possibility of a low treatment contrast, high sample attrition, and sample weighting methodology. In these instances, Mike Puma Associates requested additional information and analyses from project partners to determine the implications of the potential issues and project partners' approach to addressing them. In the discussion of sample attrition, the report from Mike Puma Associates acknowledged the need for a neutral standard (external to the project) to determine whether rate of attrition is "too high" and referenced the Institute of Education Sciences' "What Works Clearinghouse."

The validator's report ultimately certified that the calculations done by DOCCS Research office were conducted in accordance with the agreed upon evaluation methodology. The validator's report also commented on the results, exploring a variety of factors that could have influenced the final results of the pilot Phase I, such as whether the target population was appropriate, whether the theory underlying the intervention was correct, whether the intervention was implemented with fidelity, whether there was a large enough contrast between the treatment and control groups and whether the study was adequately powered. The validator notes that 45 percent of the treatment group members had their parole revoked during the Phase I observation period (returning them to prison), and thus limiting their ability to take full advantage of the CEO services.

In addition, the validator noted some striking differences between the CEO sites in New York City and Rochester, with higher rates of participation and completion in all categories observed in Rochester. While the validator does not explain these differences, except to say that the two sites offered different contexts, the report did note that this difference should be investigated further. (Puma 2017, 14) The validator also commented on the results noting that some involved with the pilot thought that the population being served faced serious challenges that could have inhibited effective engagement with CEO's services.

The validator also noted that it would be useful to assess whether the theory and assumptions underlying the intervention were correct. While CEO's intervention had been involved in a prior RCT study that showed positive impacts for high-risk recently released offenders, the validator stated that this does not "guarantee success with this study's particular sample, context, location and timing but it does provide some important evidence of the program's ability to produce positive results, at least in the instance previously studied." (Puma 2017, 19)

### 5.3 Evaluation-Validation Reports to DOL and Payment Decisions

Both PFS pilots predicted potential cost savings in their proposals based on expected reductions in bed days in prison and increased employment and earnings.<sup>17</sup> At the conclusion of Phase I, the DOL-supported period of the PFS demonstrations, both grantees submitted the required validator reports on

---

<sup>17</sup> The calculations used to determine cost savings are described in Chapter 7 of the PFS Interim Report (Abt 2016).

the evaluation findings to DOL. Because neither evaluation found projected impacts on recidivism or employment, neither grantee requested payments from DOL. In this section, we describe the reports and note any concerns the two projects expressed about the evaluations and/or the results.

### 5.3.1 Massachusetts

At the conclusion of Phase I, in April of 2017, the Urban Institute prepared a 12-page memo summarizing evaluation findings, which the State of Massachusetts shared with DOL indicating that because “current estimates do not find gains in employment or bed days avoided, there is no estimated payment to [the Massachusetts] DOL.” The memo stated that the estimates were generated based on the project’s evaluation plan and a January 2017 regression specification memo (also from Urban Institute). The memo included a description of the sample and observation period, estimation results on incarceration days and employment and payment determinations. It also included appendices with descriptive information and regression results. The estimation results included both Intent-to-Treat (ITT) estimates and Treatment-on-the-Treated (TOT) estimates, and reported the size of the effect, the standard error and 95 percent confidence interval for each. The memo did not report the means for each group on the outcomes of interest or the p-values<sup>18</sup> for the observed differences. The memo also did not report on the engagement in service outcome.

PFS partners reported that outcomes were not achieved in Phase I of the project because there was not a large enough sample observed for a long enough period of time to detect the effects of Roca on employment and recidivism. Partners reported that they realized early in the life of the project that issues with referrals were likely to put the DOL-funded phase of the project at risk. However, given that only the state of Massachusetts would benefit by receiving funds from DOL for the outcomes being met in Phase I, partners were more focused on ensuring that outcomes would be met in Phase II. As noted in Chapter 4, in fall of 2016, partners agreed to reduce the number of referrals needed and extend the enrollment period. The evaluation continued beyond the end of Phase II of the PFS pilot with a grant from the Laura and John Arnold Foundation. While these additional funds will not affect payments made to investors under the PFS pilot, they will allow the evaluation to capture the results of extending the enrollment period.

### 5.3.2 New York

The state of New York presented the findings for Phase I in a series of tables developed by DOCCS Research offices and which included the estimates of the employment, recidivism and transitional employment impacts, information on the ITT and TOT models, and the scaling factor. The tables were presented without additional discussion.<sup>19</sup> As in the case of Massachusetts, New York State did not request payment from DOL based on the impact findings, since the intervention failed to achieve its targets for reduced recidivism and increased employment and earnings. Nevertheless, as is the case in Massachusetts, New York planned to continue to evaluate the intervention through Phase II.

---

<sup>18</sup> This memo was included as part of the Massachusetts PFS Validator’s Report.

<sup>19</sup> The tables were included as part of the New York State PFS Validator’s Report.



### 5.4 Summary

Neither pilot found significant positive impacts as a result of the intervention for the first phase of the PFS projects. Both pilots continue with Phase II and hope to find positive impacts with increased sample size and a longer follow-up period.

## **6. Key Findings and Suggestions for Future PFS Pilots**

This report documents the implementation of the first phase of the PFS pilots in Massachusetts and New York. Specifically, the report examines the period of implementation after the projects were established, conducted random assignments, addressed a number of implementation challenges, and reported the results from Phase I to DOL. While the pilots did not reach the target impacts to trigger payments from DOL, the experiences of the two PFS pilots offer a range of observations relevant to other PFS efforts, to DOL and to the field in general. As noted in Chapter 1, DOL’s primary objectives in awarding the PFS grants were twofold: to test the model and understand its feasibility. With these objectives in mind, we report a number of findings and suggestions for others interested in the PFS approach.

### **Partners pointed to many positive aspects of the pilots and were optimistic about achieving outcome targets in the second phase, in spite of not achieving target outcomes in the first phase.**

Launching the PFS pilots was complex and required coordination among many partners bringing different interests and experiences to the projects. At the time of their launch, few PFS projects had been launched and so many procedures and agreements needed to be established for the first time. Both pilots successfully brought on required partners, implemented the service model and conducted rigorous evaluations. Neither project reached the planned-for outcome targets, so they did not request payment from DOL.

Partners reported that in spite of the results from the first phase, the experience of designing and implementing the PFS pilot was an important learning experience that had effects beyond the operation of the pilots. The Massachusetts state partner observed that the pilot helped shift state government staff’s thinking about the value of paying for program outcomes (instead of only funding program operations) and encouraged new ways of thinking about solving critical social problems.

### **It was challenging to conduct the projects within the four-year timeframe of the DOL grants.**

USDOL’s grants were awarded for four years, during which time the projects needed to establish agreements with partners, launch the service intervention, serve participants and measure outcomes. Both projects elected to build in a second phase, reflecting the concern that they would not be able to detect impacts within the four-year period of the grants. In particular, Roca’s full intervention takes four years, meaning there was not much time to observe outcomes for most participants.

Several of the partners in both states indicated that they viewed the first phase as “interim” and that the state-funded second phase would be a better test of the intervention’s effects. They also noted that with a longer time frame it might have been possible to do a pilot phase for which outcomes would not be included in the formal evaluation. For the Massachusetts pilot in particular, interviewees reported that a “ramp up” period would have helped to resolve a number of initial implementation issues. For government agencies with strict funding cycles, it may be important to consider what types of interventions and outcomes are appropriate in order to be able to detect impacts within the given timeframe.

The RCT evaluations were successfully implemented and approved by their respective validators, demonstrating the feasibility of their use in the PFS context. The two DOL PFS pilots were among the first Pay for Success projects in the U.S. to use an RCT design. As described in Chapter 1, other

---

## KEY FINDINGS AND SUGGESTIONS FOR FUTURE PFS PILOTS

---

PFS projects, such as those in Utah and Chicago, have faced questions about the validity of their stated outcomes and whether they warranted the payments the government made to private investors. This highlights the need for project partners and other stakeholders to have confidence in the methods used to calculate whether payments should be made. RCTs are considered the most rigorous research design, and both DOL pilots opted to use them in order to ensure confidence in stated results. DOL also specified the use of an independent validator, taking another step to ensure confidence in study results.

**While the projects met the requirement to conduct an RCT, this was one of the most difficult aspects of the PFS approach to implement.**

The use of RCTs is not without challenges, however, as they are complex to implement appropriately and require expertise and experience in putting this type of design in place in a program environment. In Massachusetts, project partners replaced the evaluator mid-way through the project, as it became clear that they needed a partner with more evaluation capacity and experience. In New York, the evaluation was conducted internally by DOCCS Research, and though strongly implemented, it created additional demands on the department. Some project partners reported that they had hoped for more ongoing performance data in order to make “course corrections” needed to address issues that arose, but this often requires a performance and data management system separate from that put in place for evaluation data collection. Some partners raised concerns about the evaluator being internal to the grantee organization and would have rather had an external evaluator. For sponsors of PFS, it is important to consider whether an external evaluator should be a formal requirement.

As the projects experienced, implementing a strong RCT requires several elements, including an adequate sample size, high levels of participation in program services among the treatment group, a strong treatment-control contrast, and high quality and accessible data. While their experiences were different, both projects had difficulties in at least some of these areas, with some addressed more effectively than others, and these factors should be considered when interpreting the results. Particularly when funding decisions are based on their results, it is important to understand the strengths and weaknesses of the evaluation design as implemented, and when the results should be interpreted cautiously.

Finally, based on the experiences of the projects, having the government sponsor identify the evaluation standards for evaluation design and analysis to be used for the evaluations of the PFS projects (such as, for example, those for the What Works Clearinghouse or DOL’s Clearinghouse for of Labor Evaluation and Research). It is important to state these standards clearly at the outset of the pilots as part of the funding announcement to ensure that all parties are aware of the needs and requirements of the evaluation. Independent review of the planned evaluation design would also ensure that standards are being followed and will be met in the final analysis of the results.

**The PFS sponsor should be specific about the required reporting on the RCT methodology and factors that may have affected the interpretation of results.**

Based on the experience the DOL pilot projects, it is important to provide detailed requirements related to how final results should be presented and what issues will be addressed. DOL made an effort to ensure confidence in the evaluation results through the use of an external validator to review the evaluation results, and both pilots implemented this partnership, and the validators approved the evaluation results. However, the differences in the content and format of both the evaluation and validation reports highlight the difficult of interpreting the results and understanding both the

---

## KEY FINDINGS AND SUGGESTIONS FOR FUTURE PFS PILOTS

---

strengths and weaknesses of the results. Evaluators for both sites made a series of analytical decisions required for impact estimation, which could have affected impact estimates. In this case, neither pilot requested payment, but if the evaluation results had shown they were closer to meeting their targeted outcomes, it is possible that there could have been disagreement among project stakeholders over evaluation methods, which could be very difficult to navigate with funding contingent upon the outcome of those disagreements.

For future PFS projects, it is worth carefully considering ways to be transparent about evaluation plans and progress through the life of the project. These methods might include requiring ongoing oversight by a validator and ongoing reporting as the evaluation progresses, updated and detailed analysis plans to be submitted after data collection has begun and evaluators are more informed about what they can expect in terms of data. Recommended methods for handling issues such as how to address low enrollment in the treatment group or inadequate or low-quality data should be detailed and clarified before the study begins.

The sponsor should clearly outline the responsibilities of and scope of work for the validator, including detailing what should be included in a final assessment report. This could include granting the validator access to the data from the evaluation and requiring specific elements in the validator's final assessment report. Specifically, this might include describing the intervention, the evaluation design and documenting how the evaluation was carried out, including major decision points, and confirm and describe the analysis methods used. If alterations were made in either the intervention or evaluation, the report should document those changes, showing how the planned designs were altered. Finally, the validator's report should present the findings clearly.

### **The DOL PFS projects required significant time and resource investment by the key partners.**

It is well known that the PFS approach is complex, involving several different partners and the development of complicated contractual agreements. Both pilots were able to successfully develop and manage the projects, but partners reiterated that doing so took a great deal of time. Staff changes were especially challenging over the course of the grants, as there was a steep learning curve for any new staff to understanding the project's technically complex concepts and design issues. The projects benefitted from pro bono support and the strong commitment at the state level to exploring the PFS approach, and it is not clear how feasible the approach would have been without such high level commitment and interest in testing the approach. As the PFS field grows, more tools and resources are being developed and knowledge shared that may facilitate the design and implementation of these projects.

### **Partners had different perspectives on the costs of the PFS pilots.**

Observations from partners highlight that costs are likely to be a consideration for stakeholders in future PFS projects (though a cost study was beyond the scope of this process study). Some partners noted that the government benefitted since it did not need to pay for services when the outcomes were not achieved. However, both states reported that they invested significant staff time in developing, managing and overseeing the project. Moreover, additional evaluation oversight, as we discussed above, would likely increase these costs, although these costs may decrease as the government becomes more familiar with the PFS approach. Some partners reported that the time required for investors, especially philanthropic investors, to oversee the projects, represented an additional cost that may be difficult for some to absorb. In contrast, for services providers, those interviewed

---

## KEY FINDINGS AND SUGGESTIONS FOR FUTURE PFS PILOTS

---

reported that the costs were not significantly different from the regular course of business, and that participating in a PFS pilot did not affect their program costs.

### **The PFS partnership structure was seen as a crucial aspect of a viable PFS project.**

Partner staff interviewed for this study consistently reported that a PFS partnership structure should include all key players that have a role in the intervention and evaluation be included in the PFS management structure. For example, in Massachusetts, the intermediary observed that a key stakeholder, the state Probation Service, was not included as a primary partner. However, Probation's role was crucial for ensuring an adequate number of referrals to the service provider. Including Probation as a primary partner, and as a signatory to the partnership agreement, would have helped to address the challenges the pilot experienced with the referral process.

For Massachusetts, the administration of the PFS pilot was placed in the state's procurement office. However, partners in this pilot observed that "housing" the PFS pilot in the agency responsible for producing the outcomes, such as the state department of labor for an employment-related intervention, may facilitate implementation of the approach. In addition, including staff with state budget knowledge is important as such staff can assist with fiscal decisions, understand the project and provide continuity across administrations.

Keeping the partners engaged for the entirety of a PFS pilot was challenging given shifting priorities and focus, especially if state administrations change. The New York pilot partners were in regular communication and understood that the pilot required regular participation from all stakeholders in order to effectively implement the evaluation. The Massachusetts state partner observed that the pilot necessitated coordination and collaboration among the partners, encouraging alignment of interests and development of innovative approaches to address the needs of the target population, calling this a "powerful feature" of the PFS approach.

### **Concluding Thoughts**

The DOL PFS pilots offer a rich set of experiences on implementing a PFS model. These experiences indicate that strong partnerships are necessary to launch and sustain the projects, and the two pilots developed and operationalized the complex partnerships needed to carry out initiatives. However, implementing a high quality, rigorous evaluation of the interventions proved to be the most difficult aspect of the PFS projects, and may have contributed to the inability to detect impacts and pay investors. This indicates the importance of carefully planning and supporting this aspect of the PFS model in future efforts.

## References

- Abt Associates, Inc. (2016). *Process Study of the U.S. Department of Labor's "Pay for Success" Pilots in Two States: Development of the Grant Applications and Initial Implementation*. Washington D.C. Report prepared for the U.S. Department of Labor, Employment and Training Administration.
- American Federation of State, County and Municipal Employees, Center for Effective Government, Economic Opportunity Institute, In the Public Interest, Keystone Research Group, Minnesota Council of Nonprofits, Oregon Center for Public Policy, & Refund America Project. (2015). *A Guide to Evaluating Pay for Success Programs and Social Impact Bonds*. Washington, D.C.: In the Public Interest.
- Bartlett, V. et al, ed. (2017). *What Matters: Investing in Results to Build Strong, Vibrant Communities*. Federal Reserve Bank of San Francisco and Nonprofit Finance Fund. San Francisco and New York.
- Berlin, G. (2016). *Learning from Experience: A Guide to Social Impact Bond Investing*. New York: MDRC.
- Blum, J., Burnett, K., Levin, M., Moazed, T., Porowski, A., & Sparks, A. (2016). *State of the Pay for Success Field II: Emerging Literature, Updates, and Tools*. Washington, D.C.: Corporation for National and Community Service, Office of Research and Evaluation.
- Disley, E., & Rubin, J. (2014). *Phase 2 Report from the Payment by Results Social Impact Bond Pilot at HMP Peterborough*. London, England (UK Ministry of Justice Analytical Series): RAND Europe.
- Disley, E., Giacomantonio, C., Kruithof, K., & Sim, M. (2015). *The Payment by Results Social Impact Bond Pilot at HMP Peterborough: Final Process Evaluation Report*. Santa Monica, C.A.: RAND Corporation.
- Eldrige, M., & Kreeger, A. (2016, May 25). *Early Results in Chicago's PFS Project Indicate Initial Success*. Washington, D.C.: Urban Institute.
- Gillespie, S., Liberman, A., Zweig, J., Hanson, D., Cunningham, M., & Pergamit, M. (2016). *Practical Considerations for Pay for Success Evaluations*. Washington, D.C.: The Urban Institute.
- Golden, M., Garvey, M., Carolan, M., & Nagendra, B. (2015). *Pay for Success in the U.S.: Summaries of Financed Projects*. Greenville, SC: Institute for Child Success.
- Harvard Kennedy School Government Performance Lab. (2017). *Social Impact Bonds 101*. Cambridge, MA: Harvard Kennedy School.
- Illinois Facilities Fund (IFF). (2017, April 28). *Chicago PFS/SIB Program – Success Payment Calculation Evaluation #2*. Chicago: IFF.
- Innocenti, M. (n.d.). *FAQs About Utah's Social Impact Bond for Early Childhood Education*. Utah: Utah.gov.

- Liebman, J. & Sellman, A. (2013). *Social Impact Bonds: A Guide for State and Local Governments*. Cambridge, MA: Harvard Kennedy School.
- McGahey, R., & Willis, M. (2017). "The Promise and Reality of Social Impact Bonds" in *What Matters: Investing in Results To Build Strong, Vibrant Communities*, edited by Bartlett, V. et al. San Francisco, C.A. and New York, N.Y.: Federal Reserve Bank of San Francisco and Nonprofit Finance Fund.
- New York State Press Release. (2012, February 17). *Governor Cuomo Announces "Work for Success" Employment Initiative for the Formerly Incarcerated*. Albany, NY.
- Notice of Availability of Funds/Solicitation of Grant Applications for Pay for Success Pilot Projects* (SGA/DFA PY-11-13). (June 2012). Washington, DC: United States Department of Labor, Employment and Training Administration.
- Notice of Funding Availability, Social Impact Partnerships to Pay for Results Act Demonstration Projects* (Funding Opportunity Number: UST-SIPPRA-2019-001). (2019). Washington, DC: United States Department of the Treasury, Office of Economic Policy.
- Pay for Success: Collaboration among Federal Agencies Would Be Helpful as Governments Explore New Financing Mechanisms*. (2015). Washington, D.C.: U.S. Office of Government Accountability.
- Pay For Success Contract among the Commonwealth of Massachusetts, Roca, Inc., and Youth Services Inc.* Dated January 7, 2104.
- Pay For Success Intermediary Agreement* [among the New York State Department of Labor, Social Finance, Inc, and Social Finance - NY State Workforce Re-Entry 2013 Manager, Inc.]. Dated October 1, 2013. Albany, N.Y.: New York State Department of Labor.
- Public Consulting Group (PCG). (April 2017). *Massachusetts Social Innovation Financing Youth Recidivism Pay-for Success Initiative: Independent Validator, Interim Outcome Estimation Validation Quarter 14 Report*. (Unpublished).
- Puma, M. (March 2017). *New York State Pay for Success (PFS) Project: "Employment to Break the Cycle of Re-Incarceration" Phase I Validation Report*. Mike Puma Associates LLC. (Unpublished).
- Propper, N. (2015, November 3). *Success Metrics Questioned in School Program Funded by Goldman*. New York: New York Times.
- Redcross, C., Millenky, M., Rudd, T., & Levshin, V. (2012). *More Than a Job: Final Results from the Evaluation of the Center for Employment Opportunities (CEO) Transitional Jobs Program*. OPRE Report 2011-18. Washington, D.C.: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Sanchez, M. (2016, May 16). *Investors Earn Max Initial Payment from Chicago's 'Social Impact Bond.'* Chicago: The Chicago Reporter.

*Second Amended and Restated Pay For Success Contract Among the Commonwealth Of Massachusetts, Roca, Inc., and Youth Services Inc.* Dated November 1, 2016.

*World's 1<sup>st</sup> Social Impact Bond Shown to Cut Reoffending and to Make Impact Investors A Return.* (2017, July 27). London: Social Finance.



## Appendix A. Pilot Interviewees

### Massachusetts Interviewees

Lead state agency: Office of Administration and Finance

State workforce agency: Executive Office of Labor and Workforce Development

Intermediary: Third Sector Capital Partners

Evaluator: Urban Institute

Independent validator: Public Consulting Group

Service provider: Roca

Investors: Living Cities, Goldman Sachs, and Laura and John Arnold Foundation

Technical assistance: Government Performance Lab at Harvard Kennedy School of Government

### New York State Interviewees

Lead state agency: NYSDOL Administrative Finance Bureau

State workforce agency: NYSDOL Division of Employment and Workforce Solutions and NYSDOL Division of Research and Statistics

Governor's Office: NYS Division of the Budget

Intermediary: Social Finance Inc.

Government evaluator: Department of Corrections and Community Service (DOCCS)

Independent validator: Mike Puma Associates, LLC

Service provider: Center for Employment Opportunities (CEO)

Investors: Living Cities and Laura and John Arnold Foundation

Underwriter: Bank of America Merrill Lynch

Technical assistance: Government Performance Lab at Harvard Kennedy School of Government

## Appendix B. Basic Concepts in Evaluation

This Appendix provides a brief overview of key concepts and terms in evaluation, including a definition of Randomized Controlled Trial evaluation design, a discussion of sample size, and a description of data sources and statistical significance.

### Randomized Controlled Trial Evaluation Design

The key to the rigor of an RCT is random assignment. In random assignment, eligible program applicants are randomly assigned, as if by lottery, either to a treatment group, which receives the intervention, or to a “control group,” which does not. Random assignment ensures that the two groups are identical in all respects both observed and unobservable, except that members of one will participate in the intervention (receive its services) and the other will not. Therefore, any differences in outcomes between these groups (e.g., different rates of employment) can be directly attributed to the intervention. RCTs are considered the “gold standard” in evaluation because this method allows programs to claim with a certain degree of confidence that changes in participant outcomes are solely because of that program.

For RCT studies, the impact estimates report the average difference in the outcomes of interest between the treatment group and the comparison group. When properly implemented, RCTs estimate the average impacts of the program under study, and these impacts can be directly attributed to the program. That is, RCT findings allow readers to say “program X *caused* outcome Y.” Well-run RCTs inform policymakers and practitioners about the effectiveness of a program and allow them to reasonably expect that similar impacts will occur again if the same program is implemented with a similar population.

### Sample Size

*Sample size* refers to the number of subjects included in the evaluation. Larger sample sizes tend to allow for more precise findings. Apart from considering the baseline sample size (i.e., the number of subjects enrolled into the study), follow-up sample sizes and the difference between the baseline and follow-up sample sizes (i.e., *attrition*) affect the precision of the findings. If a study experiences high drop-off (more than 30 percent or so) between baseline and follow-up, the findings could become less precise, and also might not be representative of the full population affected by the intervention. Additionally, evaluators should assess whether those subjects for whom they have follow-up data are systematically different from those for whom they do not have follow-up data. If the groups are different, again, the study results might not be representative of the full population affected by the intervention.

Finally, for participant services interventions, the *participation rate* is another important factor. Participation rate is defined as the number of treatment group members who participated in the intervention. If only a small portion of the treatment group baseline sample participates in the intervention, then it would be difficult to conclude that any changes in outcomes are due to the intervention, no matter the evaluation design.

## Data Sources and Follow-Up Periods

Evaluation results might be based on survey data, program data, national administrative data (such as Unemployment Insurance, UI, data), or other data sources. Regardless of source, three aspects of its data should be considered:

Validity—Do the data actually measure the outcome as intended/as it is defined?

Reliability—Do the data consistently measure the outcome?

Sensitivity—To what extent do the data measure the nuance of the outcome?

Each data source has pros and cons regarding each factor. Evaluation reports should identify the data source(s) on which the findings are based, and they should provide information about these sources.

The *follow-up period* is an important consideration not only in interpreting outcomes, but also in understanding the findings. A follow-up period should be sensible based on the timeline of the intervention. For example, if a training program lasts for 12 months, then measuring its employment effects at six months after study enrollment would be an unfair test of the program. Conversely, if a program lasts 12 months, measuring employment at 18 months could be a fair test and reasonable time frame in which to expect any employment effects to be detectable.

Some data sources are more reliable than others—for example, UI data are collected in the same manner for all individuals within a state and the data are unbiased. Conversely, program administrative data might not be collected consistently across program participants, sites, or grantees; might be incomplete; or might have human data-entry errors.

## Statistical Significance

There are several common standards for judging statistical significance—that is, for judging the strength of the evidence that the observed difference between the treatment group (pretest) value and comparison group (posttest) value is a true difference and not the result of chance. The smaller the *p*-value, the stronger the evidence. In many studies, tests are considered statistically significant if the *p*-value is less than or equal to .10 (10 percent). It and smaller *p*-values are often separately flagged by the placement of asterisks next to the estimate, as follows:

\* for .10 (10 percent likelihood of the result being due to chance)

\*\* for .05 (5 percent likelihood)

\*\*\* for .01 (1 percent likelihood)

If an estimate is not statistically significant, then the finding cannot be interpreted as the treatment group having fared any better (or worse) than the comparison group. The two groups or time frames being compared are statistically equivalent if the estimate of their difference is not statistically significant.

A lack of statistically significant findings in an evaluation can be due to many reasons. Non-statistically significant results might not necessarily mean that the program was ineffective; rather, a lack of findings could be due to evaluation limitations, such as too short a follow-up time period, not enough sample members to detect an impact, poor measurement techniques, or incorrect analyses. Or

lack of effects could be due to a failure to implement the program as designed or to contextual factors that program operators had no control over, such as changed laws or policies, administration change, new competing similar programs, or changes in the local economy.

Lack of effects, however, could mean that the program itself was not effective. The program context and implementation study findings can help explain why or why not effects were detected. Quantitative findings should not be viewed in a vacuum, but instead in connection with all other aspects of the program under study.

## Appendix C. Recommendations for PFS Validator Reports

Abt Associates reviewed the Validators' Reports for the two Pay for Success (PFS) DOL projects and found that it was difficult to link the reports back to the planned analysis and, in some cases, to find evidence to support key observations. Below are recommendations for future Evaluation or Validator's Reports for PFS projects, particularly those that have evaluations with a Randomized Controlled Trial (RCT) design or quasi-experimental designs employing statistically matched comparison groups. Overall, the reports should be "stand alone" documents that thoroughly describe the intervention, evaluation, analysis, and final outcomes. They should also include the information on the following:

**Study Sample:** The report should include a section with information on recruitment, eligibility criteria for inclusion in the sample and the random assignment process, i.e., how the groups were formed. The report should describe how the enrollment of the sample (for random assignment, enrollment in services) compared to the number that was targeted at the outset of the intervention/evaluation. If there is variation, the report should clearly document and describe what the final sample size means for estimating outcomes in terms of power. The report should also document and describe how eligibility of the sample was verified.

**Random Assignment Process:** The report also should describe the implementation of the random assignment process. While this process is described in the PFS pilots' grant agreements, the report should present the process to ensure the reader can understand what happened, as well as to document if there were changes from the originally proposed design. The report should also describe how integrity was maintained, i.e., how the process in assigning participants to treatment and control groups was checked and verified. Finally, the report should describe how the referral process operated and if the random assignment process yielded equivalent groups on observable characteristics.

**Back-stop Methodology:** For a PFS pilot that includes a "back-stop" methodology, the report should describe that methodology, when and why it can be or was invoked, the steps the evaluator went through to undertake this process, how it was described in the grant agreement, if it was changed, and what were the changes and reasons for them.

**Data Sources, Outcome Measures and Timing of Data Collection:** A brief description should be included on the data sources used to construct study variables, as well as on data elements, outcome measures, data collection methods and timing.

**Outcomes/Impacts and Analytic Methods:** Clear and easily understood information on results should be presented in the report, and include information on outcomes and impacts, analytic methods (e.g. the Greek analysis models) used to analyze each outcome or address each research question, weighting and adjustment, sample sizes, and data sources. A sample table is provided below.

Outcome	Adjusted Treatment Percent	Unadjusted Control Percent	Adjusted Impact Pp	Standard Error	p-Value	Effect Size

Note: Sample sizes were ## treatment and ## control.

**Treatment of Missing Data and Baseline Equivalence:** If missing data are expected in the underlying data used to construct study measures, then the methods used to address missing data need to be reported. For example, the dummy variable method can be used to address missing baseline variables, and case deletion can be used to address missing outcomes. For reporting purposes, overall and differential attrition should be calculated for each outcome to help assess if the RCT is broken, i.e., if there is so high a rate of missing treatment and control group data that the two may no longer be similar on all covariates. The report should address if there are missing data rates for any outcome, and whether the attrition rate exceeds the accepted attrition standards for an RCT (as defined, for example, in DOL’s Clearinghouse for Labor Evaluation and Research) for baseline equivalence in characteristics across the treatment and control groups.

**Information on Participation in Treatment Group Services:** The report should show the raw numbers and calculated rates of noncompliance with random assignment. Noncompliance with random assignment could occur if a participant in the treatment group did not receive the intervention (no-show) or if a participant in the control group did receive the intervention (cross-over).

**Sensitivity Analyses (if applicable):** Sensitivity analyses can be conducted to examine the robustness of the findings about impacts on the planned outcomes. Examples of this type of analysis can include (i) sensitivity of the findings to specific analytic decisions made about the model specification. In particular, impacts can be re-estimated using other model specification, e.g., using logistic regression instead of linear regression, including (or not including) covariates in the analysis model, or using treatment on the treated (TOT) estimator instead of an intent to treat (ITT) estimate; (ii) sensitivity of the findings to specific methods used for dealing with missing covariates, e.g., using case deletion for missing covariates or a dummy variable method.

**Accessible Language:** It would be helpful if the report were written so as to be as accessible as possible to a general non-technical audience. Based on the SGA, DOL anticipated that launching and implementing a PFS project would be a complex endeavor, and that the pilots would be of interest in an emerging field and to state, local and tribal governments interested in using the approach.