Next Generation of Enhanced Employment Strategies Project: Evaluation Design Report

OPRE 2024-154

August 2024

NEXT GEN

This page has been left blank for double-sided copying.

Next Generation of Enhanced Employment Strategies Project: Evaluation Design Report

OPRE Report 2024-154
August 2024

April Wu, Kristen Joyce, Annalisa Mastri, Sheena McConnell, Jody Schimmel Hyde, and Jacqueline Kauff

Submitted to:

Project Officers: Hilary Bruck, Marie Lawrence, Sarita Barton, and Megan Reid
Office of Planning, Research, and Evaluation
Administration of Children and Families
U.S. Department of Health and Human Services

Contract Number: HHSP233201500035I/HHSP23337035T

Submitted by:
Project Director: Annalisa Mastri
Mathematica
1100 First Street, NE, 12th Floor
Washington, DC 20002-4221
Reference Number: 50718

Suggested citation:
April Wu, Kristen Joyce, Annalisa Mastri, Sheena McConnell, Jody Schimmel Hyde, and Jacqueline Kauff. (2024). Next Generation of Enhanced Employment Strategies Project: Evaluation Design Report. OPRE Report #2024-154. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Disclaimer:
The views expressed in this publication do not necessarily reflect the views or policies of the Office of Planning, Research, and Evaluation, the Administration for Children and Families, or the U.S. Department of Health and Human Services.

This report and other reports sponsored by the Office of Planning, Research, and Evaluation are available at www.acf.hhs.gov/opre.

This page has been left blank for double-sided copying.

Contents

Exhibits

Executive Summary

Overview

Many people face complex challenges to obtaining, retaining, and advancing in employment, that in turn affect their ability to be economically independent. Although current research suggests some promising strategies for people facing complex challenges, there is still much to learn about how to best serve this population.

To identify and study innovative employment programs for people facing complex employment challenges, the Office of Planning, Research, and Evaluation (OPRE) within the Administration for Children and Families, U.S. Department of Health and Human Services is sponsoring, and Mathematica is conducting, the Next Generation of Enhanced Employment Strategies (NextGen) Project. The NextGen Project is part of OPRE's Innovative Strategies for Addressing Employment Barriers Portfolio, which seeks to rigorously evaluate the "next generation" of employment strategies for individuals with low incomes, and is partnering with the Social Security Administration (SSA) on select evaluations.

This report describes the design of the NextGen Project, which is evaluating the effectiveness of four programs that provide services for people with physical, mental, or emotional health challenges.[1] Many of these people are potential applicants for Supplemental Security Income (SSI). The results of the NextGen Project will inform policymakers and practitioners interested in helping people who face complex challenges to employment become economically secure. The findings are also intended to help SSA better understand the types of programs that can connect or reconnect potential SSI applicants to work before they apply for benefits. The key research questions include the following:

- **How were programs implemented?**
  - What is the context in which each program was implemented?
  - What is the design of each program?
  - How and how well was each program implemented?
- **What is the cost to implement the program?**
  - What is the total cost per participant per month for each program?
  - What is the total cost per participant for each program?
- **Do the programs improve outcomes?**
  - Does each program affect participants' employment outcomes?
  - Does each program affect participants' economic independence?
  - Does each program affect the amounts and types of services participants receive?

[1] The NextGen Project includes an impact evaluation of a fifth program, Work Success, which serves Temporary Assistance for Needy Families (TANF) recipients and other job seekers who use American Job Centers in Utah. The Work Success design report is available here https://www.acf.hhs.gov/opre/report/design-and-analysis-plan-impact-study-work-success. Implementation and impact reports are forthcoming.

- Does each program affect participants' health and other outcomes?

To address these research questions, each of the four evaluations includes descriptive, cost, and impact studies, described further in the following sections.

Programs included in the evaluation

The NextGen team identified and selected programs for evaluation that would build evidence about how best to serve people facing complex challenges to employment. The NextGen team identified four programs that met three general criteria: (1) they address OPRE's and SSA's research interests; (2) they were well implemented or could be with some assistance; and (3) rigorously evaluating them was feasible or could be with some assistance. For each initially selected program, the team conducted an evaluability assessment.

The team selected four programs:

- **Bridges from School to Work (Bridges)** serves young adults (17–24) with disabilities who are transitioning out of high school. The program aims to meet the needs of both the young adults it serves and local employers. Bridges staff use a strengths-based approach focused on the young adults' skills, interests, and abilities rather than a deficit-oriented framing focused on their disabilities. The evaluation is taking place in six cities.
- **Individual Placement and Support for Adults with Justice Involvement (IPS-AJI)** provides assistance to adults with mental health issues who are reentering the community after incarceration or who have received an alternative sentence. The program offers participants mental health treatment and employment assistance at mental health centers using the IPS model. The NextGen Project is testing IPS-AJI in five mental health centers, located in Claremore, Oklahoma; Florence, South Carolina; Memphis, Tennessee; Moline, Illinois; and Norman, Oklahoma.
- **Philadelphia Workforce Inclusion Networks (Philly WINs)** serves adults with low incomes and chronic physical, mental, or emotional conditions or disabilities that could limit their employment. The program develops relationships with and provides technical assistance to a network of employers that provide inclusive workplaces and prepares program participants for jobs at these and other employers. Philly WINs is located in Philadelphia.
- **Western Mass Mental Health Outreach for MotherS Partnership℠ (Western Mass MOMS)** serves adult caregivers who identify as women or nonbinary, have low incomes, and exhibit depressive symptoms. The program is based on the MOMS Partnership® model, which is designed to reduce depressive symptoms, improve social connections, and promote economic well-being among mothers. For Western Mass MOMS, the NextGen team worked with the MOMS Partnership model developers and the service provider to add employment services. Western Mass MOMS is located in Springfield and Holyoke, Massachusetts.

The NextGen team conducted activities to prepare these four programs for the evaluation. This work involved identifying each program's core components; providing technical assistance on implementing the evaluation and, in some cases, program services; and conducting formative evaluations of some programs.

Descriptive study

Each descriptive study documents the programs and their operations. The main objectives of the descriptive study are: (1) to provide information useful to other organizations that might consider implementing similar programs, and (2) to interpret the impact findings. To meet these objectives, the descriptive study for each program discusses the following:

1. **The context in which the program was implemented.** A program's effectiveness is influenced by the context in which it is implemented. Context includes the characteristics of the population of interest, the organization implementing the program, the partners involved in providing the services, and the local community at the time of the evaluation.
2. **The design of the program.** The study documents in detail how each program being evaluated is intended to be implemented. The descriptive study provides information other organizations might want to consider in deciding whether they could implement a similar program. It identifies the core components and any principles that guide the program's implementation; other aspects of the program that might be important but were not deemed core components; the program's logic model; and the criteria for program eligibility.
3. **The program's implementation.** The descriptive study assesses whether the program was implemented as designed. The assessment includes a description of how the implementers adapted the program design to fit the local context and environment.
4. **Potential implications for the impact study.** The descriptive study includes discussion of factors that might contribute to impacts observed through the impact study as well as factors that might inhibit the evaluation's ability to find impacts. For instance, programs being implemented as designed and participants receiving the intended dosage of service tend to support impacts; programs being too similar to other services in the community or low service take-up tend to dampen impacts.

The descriptive studies follow five main principles to meet the study objectives and answer the research questions. Each descriptive study (1) is guided by a conceptual framework; (2) uses a focused approach to data collection grounded in the program's core components; (3) relies on data from multiple sources, collected over time; (4) incorporates perspectives of program staff and participants in data collection and analysis; and (5) follows a structured approach to analysis.

The NextGen team collected data from numerous sources for the descriptive study. Primary data sources for the report include the following:

- Ongoing conversations with program leaders and staff as part of technical assistance implementing the study
- Interviews with program leaders and staff
- Interviews with partners and employers
- Staff and leadership surveys
- Program observations and job shadowing

- Demographic, economic, and background information on participants taken from a baseline survey at study enrollment and follow-up surveys
- In-depth interviews with participants
- Third-party fidelity reviews
- Service receipt data from Random Assignment, Participant Tracking, Enrollment, and Reporting (RAPTER®) or programs' management information systems
- Program documents, literature, and data

The NextGen team analyzed these qualitative and quantitative data using descriptive, comparative, and thematic analysis approaches.

Cost study

The NextGen team is estimating the cost of each program—both overall and per participant. As part of this work, the team is collecting financial data from program managers using a customized Microsoft Excel workbook that reflects each program's operations and the structure of its existing data. The team will use data collected from RAPTER or programs' own management information systems to determine the number of study participants and the duration of the services they receive.

The NextGen team is using the "ingredients" approach to estimate the total cost of each program (National Academies of Sciences, Engineering, and Medicine 2016; Levin and Belfield 2015). Under this approach, the team (1) lists each resource required to deliver a program, (2) determines the monetary value of each, and (3) sums these values to estimate total annual costs. The team values all resources—labor, facilities, equipment, and overhead—required to deliver the program, including those that may not appear on an organization's expenditure records (for example, volunteers and other in-kind resources). The team is estimating the costs of the evaluation and subtracting them from the total costs.

In addition, the cost study team is estimating the components of total costs, such as labor and overhead, and the program's funding sources.

Impact study

The impact studies seek to determine each program's effectiveness in helping people who face complex challenges to employment become economically independent. The following key research questions address the extent to which each program improves outcomes of interest:

- **Does each program affect participants' employment outcomes?** The NextGen team will examine outcomes such as earnings, employment, job retention, and job quality.
- **Does each program affect participants' economic independence?** The NextGen team will examine whether each program reduces the need for SSI, per SSA's interest in better understanding the types of programs that effectively connect or reconnect potential SSI applicants to work before they apply for benefits. The team will examine whether some of

those programs reduce the need for Temporary Assistance for Needy Families and the Supplemental Nutrition Assistance Program.

- **Does each program affect the amounts and types of services participants receive?** The NextGen team will examine whether each program increases receipt of employment services. The team will examine whether some of those programs increase the use of mental and physical health services.
- **Does each program affect participants' health and other outcomes?** The NextGen team will examine whether some of the programs improve participants' mental health conditions, degree of social support, involvement with the criminal justice system, and other outcomes.

In addition to these key research questions, the impact studies will address the following research questions to shed light on differences in impacts by service receipt and participant characteristics:

- Are programs more effective for some groups of participants than others?
- To what extent do impacts on shorter-term outcomes such as mental health and social support explain the longer-term impacts on employment?
- Are programs effective for study participants who receive different amounts of program services?

People who are eligible for program services and have consented to participate in the study will be randomly assigned to (1) a program group that is offered the program services or (2) a comparison group that is not offered program services but is free to seek other services available in the community. The four programs vary in size, but the sample size at each program needs to be large enough to detect impacts that would be expected for the program. With these considerations in mind, the target sample size for all programs was set at 1,000 study participants—500 each in the program and comparison groups. The first program started enrollment in June 2021, and the last program started in May 2022. Enrollment will end for each program in June 2024.

The NextGen team will collect data from the following data sources to support the impact study:

- **Baseline survey.** All study participants will complete a baseline survey at the time of random assignment. The NextGen team will use these data to describe the characteristics of study participants, check that random assignment has created program and comparison groups with similar characteristics, control for baseline characteristics when estimating program impacts, construct weights to adjust for survey nonresponse, support subgroup and other analysis, and locate study participants for follow-up surveys.
- **Follow-up surveys.** The NextGen team will collect data through two follow-up surveys of study participants. The first follow-up survey will be administered to study participants six or nine months after random assignment (depending on the program); the second follow-up survey will be administered 12 months after the first one. The NextGen team will

use these data primarily to assess program effectiveness and examine the mechanisms through which program impacts operate.

- **Administrative records.** The NextGen team will collect data on quarterly employment and earnings, dates of new hire, and unemployment insurance benefit receipt using the National Directory of New Hires, a database maintained by ACF's Office of Child Support Services. The team will also collect data on earnings and disability benefit receipt using several SSA data sources, including the Master Earnings File. For select evaluations, the team will collect administrative data on other public benefits and involvement with the criminal justice system.

The main impact estimates for all outcomes will be based on the evaluation's experimental design. With random assignment, the members of the program and comparison groups should be, on average, similar in characteristics at the time of study enrollment. Our basic analytic approach is to compare the mean outcomes of members of the program and comparison groups after study enrollment. This approach will provide unbiased estimates of the impacts of the program. However, to increase the precision of the impact estimates, the study team will use a linear regression model to control for differences in the baseline characteristics of the program and comparison groups.

1. Introduction

Many people face complex challenges in obtaining, retaining, and advancing in employment, which in turn affects their ability to be economically independent. Although the existing research evidence suggests some promising strategies for people facing complex challenges, there is still much to learn about how to best serve this population.

This report describes the design of the evaluations of four programs designed to assist people who face complex challenges to employment become economically independent. The evaluations are being conducted under the Next Generation of Enhanced Employment Strategies (NextGen) Project, funded by the Office of Planning, Research, and Evaluation (OPRE) within the Administration for Children and Families (ACF) in partnership with the Social Security Administration (SSA).[2,3]

The NextGen Project is part of OPRE's Innovative Strategies for Addressing Employment Barriers Portfolio, which seeks to rigorously evaluate the "next generation" of employment strategies for individuals with low incomes. As part of this portfolio, OPRE is partnering with SSA to incorporate a focus on employment-related early interventions for individuals with current or foreseeable disabilities who have limited work history and are potential applicants for Supplemental Security Income. SSA is providing financial and technical support for the evaluation and/or service provision for select interventions within the NextGen Project.

Need for more information about promising interventions

People face challenges to employment for many complex and often interrelated individual and structural reasons (Hong et al. 2022). Individual challenges include: physical and mental health conditions; substance use disorder; early childhood trauma; intimate partner violence; criminal justice system involvement; and lack of education, credentials, or work experience (Avellar et al. 2018; Treskon 2016). Structural challenges include: lack of jobs; discrimination; lack of high-quality education and training programs; immigration and criminal justice system policies; and lack of affordable childcare, housing, public transportation, and health services. Structural challenges can intersect with individual challenges. For example, good quality education may not be available for some people of color (Hong et al. 2022). Moreover, the stresses and uncertainty of a lack of income can be overwhelming, leaving less mental bandwidth for effective development and use of self-regulation skills—skills needed to finish tasks, stay organized, and control emotions—that are critical for attaining, keeping and advancing in a job (Mullainathan and Shafir 2013).

Despite the need for interventions for people facing complex challenges to employment, the evidence about effective employment interventions based on two decades of research conducted by ACF, SSA, the U.S. Department of Labor (DOL), and others is far from conclusive. A recent meta-analysis of the effectiveness of employment interventions for

[2] The project also includes an impact and descriptive evaluation of Work Success, a coaching program, and a descriptive study of the Wellness, Comprehensive Assessment, Rehabilitation, and Employment (WeCARE) program. The designs of these evaluations are summarized in Wu et al. 2024 and Sattar et al. 2022, respectively.

[3] The project is being conducted under the Office of Management and Budget control number 0970-0545.

populations with low incomes found that many employment programs have a positive impact, but the magnitude of impacts on earnings are small—typically not enough to lift people and their families out of poverty (Vollmer et al. 2017). In addition, the impacts of programs often are short lived. For example, of 13 subsidized employment programs studied in a rigorous evaluation sponsored by ACF and DOL, most improved employment and earnings in the first year after study enrollment when program participants were receiving the subsidy, and about half maintained those impacts through the second year, but only four had impacts beyond two years when the jobs were no longer being subsidized (Cummings and Bloom 2020). Moreover, some people facing complex challenges to employment, such as low levels of education or criminal justice system involvement, are not eligible for programs found effective for more advantaged people (Peck et al. 2018).

Launch of the NextGen Project to address this need

The NextGen Project builds on findings and lessons from past and ongoing evaluations by identifying and rigorously evaluating the "next generation" of employment strategies for populations with complex challenges to obtaining, retaining, and advancing in employment. The four programs being evaluated in the NextGen Project all provide early interventions for people with current or foreseeable disabilities who have limited work history and are potential applicants for Supplemental Security Income (SSI). Each of the four evaluations is designed to: (1) describe the program's design and implementation through a descriptive study; (2) estimate the program's cost per participant via a cost study; and (3) provide evidence about the program's effectiveness through an impact analysis.

The results of these evaluations will inform policymakers and practitioners interested in helping people facing complex challenges to employment become economically independent. The findings are also intended to assist SSA in better understanding the types of programs that effectively connect or reconnect potential SSI applicants to work before they apply for the benefits.

OPRE sponsors another project—Building Evidence on Employment Strategies for Low-Income Families (BEES)—which also has a goal of increasing the field's understanding of the types of programs that can improve labor market outcomes for people with complex challenges to employment. The NextGen Project has a special focus on the role of market-oriented programs, in which employers play a key role. BEES has a special focus on the role of programs that serve adults whose employment prospects have been affected by substance use disorders. BEES and the NextGen Project are coordinating on the implementation of these studies.

Overview of programs to be evaluated

The four programs that the NextGen Project is evaluating include:

1. **Bridges from School to Work (Bridges)** serves young adults (17–24) with disabilities who are transitioning out of high school. The program has been operating for 30 years and is currently located in 12 urban areas across the United States. It aims to meet the needs of both the young adults it serves and local employers. Bridges staff use a strengths-based approach focused on the young adults' skills, interests, and abilities rather than a deficit-oriented framing focused on their disabilities. Each participating young adult works with

a Bridges staff member who provides intensive one-on-one support in helping that young adult become ready for employment. Staff members also assist young adults during their job search and provide ongoing support for up to one year after they are placed in a job. The evaluation is occurring in six cities.

2. **Individual Placement and Support for Adults with Justice Involvement (IPS-AJI)** serves people with mental health issues, not necessarily serious, who are reentering the community after incarceration or who have received a sentence through a mental health or drug court that does not involve incarceration. The program offers participants mental health treatment and employment assistance at mental health centers using the Individual Placement and Support (IPS) model, which has been shown to be effective for people with serious mental health issues. Key principles of the IPS model include rapid search for a job that matches the participant's interests and continuing, time-unlimited support while the participant is employed. IPS staff learn about employers' needs and match participants to job opportunities accordingly. IPS participants can receive mental health treatment, and the mental health providers meet regularly with the employment service providers to discuss participants' progress and challenges. For adults with justice involvement, employment service providers also support participants in communicating their history of justice involvement to employers. The program also offers participants help in understanding the public assistance benefits available to them. The NextGen Project is testing IPS-AJI in six mental health centers, located in Claremore, Oklahoma; Florence, South Carolina; Memphis, Tennessee; Moline, Illinois; Norman, Oklahoma; and Oklahoma City, Oklahoma.
3. **Philadelphia Workforce Inclusion Networks (Philly WINs)** serves adults with low incomes and chronic physical, mental, or emotional conditions or disabilities that could limit their employment. The program develops relationships with and provides technical assistance to a network of employers that provide inclusive workplaces. These workplaces are intended to allow workers to be productive and feel welcomed, and to accommodate their needs. Eligible participants are identified at four Pennsylvania CareerLink© centers (which are part of the American Job Center network) in Philadelphia. Services the program offers to participants include assessing their interests and capabilities, matching them with employment opportunities at those employers with inclusive workplaces, providing accommodations and other services to support their job search efforts, supporting their integration into the workforce, and providing follow-up services as needed at the job site or in the community.
4. **Western Mass Mental Health Outreach for MotherS Partnership℠ (Western Mass MOMS)** serves adult caregivers who identify as women or nonbinary, have low incomes, and exhibit depressive symptoms. The program is based on the MOMS Partnership® model—a program designed to reduce depressive symptoms, improve social connections, and promote economic well-being among mothers. The core of the MOMS Partnership is a series of eight 90-minute classes based on cognitive behavioral therapy principles. The classes are designed to help participants manage their stress, better communicate with others, connect with their children, and work toward their goals. The class is facilitated by a clinician and a staff member with lived experiences similar to those of the participants. The program also connects participants to needed supports and offers a financial incentive for attending classes. For Western Mass MOMS, the NextGen team worked with the MOMS Partnership developers and the service provider to add employment services. Employment services are offered both in groups—called Moving

Forward groups—and, if needed, in one-on-one meetings with an employment specialist. The groups and the one-on-one meetings could include activities such as networking, resume development, job search, interviewing, or issues that arise on the job. Western Mass MOMS is located in Springfield and Holyoke, Massachusetts.

Exhibit 1 summarizes the main features of the programs being evaluated in the NextGen Project and the timing of their impact studies.

Exhibit 1. Summary of programs being evaluated under the NextGen Project

Program name	Focal population	Location	Key components of program	Impact study start date
Bridges	Young adults with disabilities transitioning out of high school	Study is occurring in a subset[4] of the following urban areas where Bridges operates: • Atlanta, GA • Baltimore, MD • Boston, MA • Chicago, IL • Dallas, TX • Fort Worth, TX • Los Angeles, CA • New York, NY • Oakland, CA • Philadelphia, PA • San Francisco, CA • Washington, DC	Focuses on needs of employers Provides intensive one-on-one employment support to young adults, from job readiness activities to support in a job for up to one year after placement	Random assignment began in August 2021
IPS-AJI	Adults with mental health issues who are reentering the community after incarceration or who have received an alternative sentence	• Claremore, OK • Florence, SC • Memphis, TN • Moline, IL • Norman, OK • Oklahoma City, OK	Assists participants in rapid search for competitive jobs Matches participants with jobs that meet their interests Coordinates between mental health treatment providers and IPS staff Develops jobs with employers Supports communicating justice involvement history to employers Provides long-term support to participants after job placement Provides benefits planning for participants	Random assignment began in July 2021

[4] Some school districts required anonymity to participate in the evaluation. The evaluation is occurring in six of these locations.

Program name	Focal population	Location	Key components of program	Impact study start date
Philly WINs	Adults with low income and chronic physical, mental, or emotional conditions or disabilities that limit their employment	Philadelphia, PA	Develops relationships with a network of employers and provides them with technical assistance about inclusive workplaces Matches participants to employers based on participants' interests and capabilities Provides accommodations and other services to support participants' job search Supports participants' integration into the workforce and provides ongoing follow-up services as needed at the job site or in the community	Random assignment began in May 2022
Western Mass MOMS	Adult caregivers who identify as women or nonbinary, have low income, and have depressive symptoms	Springfield, MA Holyoke, MA	Provides a series of eight classes on managing stress, led by a clinician and a person with relevant lived experience Connects participants to employment services Connects participants to supports in the community Provides financial incentives to participants for attending classes	Random assignment began in March 2022

More information about these programs is provided in Appendices A-D and on OPRE's website.

Overview of study design

Each NextGen Project evaluation is distinct but has important common elements. The impact studies are all randomized controlled trials (RCTs). For the impact studies, we will collect survey data at a minimum of three points in time (baseline and two follow-ups) and collect administrative data. Each evaluation also includes a detailed descriptive study and a cost study. Efforts have been made to infuse culturally responsive and equitable evaluation principles throughout the project. Because of the differences across focal population and program design, the exact research questions differ and some of the data collected differs. Hence, data will not be pooled across evaluations—the findings of each evaluation will be analyzed and reported on separately. However, the overall findings from the four evaluations will be synthesized at the end of the project.

Impact study

The goal of each impact study is to assess the program's effectiveness. People eligible for each program and who consent to be in the study are randomly assigned to either a program group or a comparison group. People in the program group are offered the program's services. Those in the comparison group are not offered the program services but can receive other services provided in the community. The random assignment is conducted via a web-based management information system called RAPTER® (Random Assignment, Participant Tracking, Enrollment, and Reporting).

The NextGen team will assess the effectiveness of programs based on differences in outcomes between members of the program and comparison groups. For all programs, we will examine impact estimates for measures of employment, earnings, and application for or receipt of SSI. Depending on the program, other outcomes will include mental health, perception of social support, receipt of public assistance, and criminal justice system involvement.

Outcomes for each study participant will be tracked by surveys and administrative data. Outcomes will be measured using data collected from two telephone surveys of study participants conducted, depending on the program, 6–9 months and 18–21 months after study enrollment, as well as administrative records covering the same period.

Descriptive study

Each descriptive study will document the program and its operations. We will assess whether programs have been implemented as designed and thereby help interpret findings from the impact studies. If programs are found effective, other programs can use descriptions of their operations to consider replicating them. The descriptive studies will draw on multiple data sources, including ongoing conversations with program staff; semi-structured interviews with program leaders and staff; staff and leadership surveys; observations; in-depth interviews with participants; data about service receipt collected by program staff; program data, documents, and literature; and study participant surveys.

Cost study

For each program, we will estimate the cost per program participant. To do so, we will use data collected from RAPTER and the management information systems the programs use, as well as financial data requested from the programs. If the programs are found to be effective, we will also conduct a benefit-cost analysis by combining these cost estimates with a monetized estimate of the benefits from each impact. Both the benefits and the costs will be assessed from the perspective of an organization replicating the program and society as a whole.

Roadmap for the rest of report

The remainder of the report describes the evaluation plans in more detail. Chapter 2 describes the process for identifying and selecting programs for evaluation, as well as preparing them for the evaluation. Chapter 3 describes the descriptive study, including the research questions addressed, the data collection strategy, and the proposed analytic

approach. Chapter 4 describes the cost analysis: its goals, the data collection strategy, and the analytic approach. Chapter 5 provides details on the design of the impact study, including research questions, random assignment, data needs and sources, and the analytic approach to estimating program impacts and benefits. A separate appendix for each evaluation provides more details on each program and the design of its evaluation (Appendices A–D).

2. Identification, selection, and preparation of programs

The NextGen team identified, selected, and prepared programs for evaluation that would build evidence about how best to serve people facing complex challenges to employment. This chapter describes the criteria for selecting programs for evaluation, the activities we undertook to identify and select programs that had those characteristics, and how we prepared the selected programs to launch the evaluations.

Selection criteria

We searched for programs that met three general criteria: (1) they address OPRE's and SSA's research interests; (2) they were well implemented, or could be with some assistance; and (3) rigorously evaluating them was feasible, or could be with some assistance.

Alignment with OPRE and SSA priorities

OPRE's research objective for the NextGen Project was to identify and evaluate innovative programs designed to promote employment and economic security among people facing complex challenges to employment. For the NextGen Project, OPRE was particularly interested in programs that involve employers. Many programs have been found effective when they involve employers in the program design, focus on their needs as well as those of the program participants, or partner with employers to provide work opportunities (McConnell et al. 2014). In addition, OPRE required that selected programs already had some evidence of their effectiveness. It did not require that this evidence be based on an RCT—it could be from a non-experimental evaluation, an outcome study conducted by the program or an external evaluator, a descriptive study focusing on the outcomes of program participants, or an analysis of program data strongly suggesting positive outcomes.

SSA's research objective was to identify and evaluate programs designed to improve the economic independence of potential SSI applicants before they applied for SSI benefits. Hence, SSA was interested in programs that primarily serve working-age individuals with current or foreseeable disabilities, little or no work history, and no recent or current application or receipt of SSI benefits.

Implementation strength

We selected programs to be evaluated under the NextGen Project that were likely to be implemented well at the time of evaluation. Otherwise, null or negative findings could not necessarily be attributed to the program but could instead be due to its poor implementation. For example, if many participants did not receive the intended dosage of services, a lack of impacts could be because they did not receive the services rather than the services being ineffective. This type of ambiguity can reduce the policy relevance of the findings.

We conceptualized implementation quality broadly. We regard a program as being implemented well if it has a well-defined, standardized model; staff have training materials and are well-prepared to deliver services; participants receive the intended amount and

sequence of services; participants engage in services; participants and staff are enthusiastic about services; program leaders have positive attitudes about the program; and facilities and infrastructure are appropriate. Some programs, such as IPS, have measures of fidelity to the model, and hence measures of implementation quality. However, many programs do not have well-defined fidelity measures; thus, we assessed implementation quality using our own framework (as described in Chapter 3).

Feasibility of evaluation

We selected programs for evaluation only if conducting an experimental evaluation of them was feasible. We based this assessment on the following criteria:

- **Random assignment is feasible and study groups can be maintained.** Program leaders and staff must be willing to conduct random assignment; creating research groups that can receive different services must be feasible; and study participants must receive only services appropriate to their study group assignment. Thus, the program must have mechanisms in place to keep comparison group members from receiving services that only program group members are supposed to receive.
- **The program can recruit a sufficient number of people for the study and offer services to all program group members.** An experimental evaluation requires that enough people who are eligible for the program and consent to participate in the study can fill both a program and a comparison group. Furthermore, the program must be able to serve all of the people assigned to the program group. Each evaluation needs to have enough study participants that we would likely detect the impact expected from the program. As discussed in Chapter 5, we assumed we would require a sample size (program and comparison group members together) of about 1,000 in each program.
- **The comparison group receives notably different services from the program group.** The services available to the comparison group serve as the counterfactual for the evaluation—serving as a comparison for program services. The contrast between services offered to the program group and those available to the comparison group—either through the organization offering the program or elsewhere in the community—must be strong enough to produce differences in participant outcomes that the study is likely able to detect.
- **The likelihood of contamination is low.** If the presence of the study influences the services the comparison group receives (known as contamination), estimated impacts will be biased. Contamination may occur, for example, when the same staff are providing one type of services to the program group and another to the comparison group. In those cases, the approach to serving the program group can influence—possibly unconsciously—the way staff provide services to comparison group members.

Activities for identifying and selecting programs

Our identification and selection activities fell under three broad categories: (1) identifying promising types of programs; (2) identifying specific programs; and (3) conducting evaluability assessments of promising programs.

These activities occurred over three years. Although we identified some promising types of programs within the first year, it took much longer to identify specific programs and assess their evaluability.

Identifying promising types of programs

To identify the types of programs to consider, we reviewed the literature on, and talked with experts about, promising programs. We focused on employment programs that serve specific populations facing multiple challenges to employment and who might consider applying for SSI (for example, people with long-term public benefit receipt, chronic health conditions or disabilities, and justice system involvement). We focused especially on programs that have employer involvement.

The literature and discussions with experts highlighted several types of programs that held promise for the NextGen Project, including the following:

- **Sector-based training models**, which use input from employers to design eligibility criteria for the program, training content, and provision of credentials (Ziegler 2015). Employers sometimes also provide participants with work-based training opportunities, such as internships or on-the-job training. An experimental study found that three sector-based training programs had strong positive impacts on earnings (Maguire et al. 2010).
- **Career pathways programs**, which allow participants to progress through education and training in multiple, discrete segments, have been found effective for short-term employment outcomes (Farrell and Martinson 2017; Gardiner et al. 2017; Glosser et al. 2017; Rolston et al. 2017; Martinson et al. 2018; Peck et al. 2018).
- **The IPS model** (described in Chapter 1) has been shown effective in many studies of those with serious mental illness (Frederick and VanderWeele 2019) but has not been tested extensively with a broader population.
- **Work experience**, paid or unpaid, has been found to be an element of some effective programs. In the Employment Retention and Advancement Study, which included evaluations of 16 programs, the program with the most lasting impacts was the Personal Roads to Individual Development and Employment (PRIDE), which provided unpaid work experience, job search, and placement services, along with services to address challenges (Butler et al. 2012). The Progressive Employment model, the main component of which is rapid involvement of the participant with an employer in paid or unpaid work experience, has shown some promise in non-experimental studies for people with disabilities (Mann et al. 2018).
- **Apprenticeships** combine structured on-the-job training with technical instruction in a specific industry or occupation. A large, non-experimental evaluation of Registered Apprenticeship programs found they were effective at improving participants' employment outcomes (Reed et al. 2012). However, apprenticeships tend to have entry requirements that exclude people facing serious employment challenges.
- **Social enterprises**, which produce goods or services in the competitive market and employ paid workers who face complex challenges to employment, also have shown positive impacts on employment, albeit using non-experimental methods. A study of

eight social enterprises in California found them to be effective in improving employment outcomes (Maxwell and Rotz 2017; Rotz et al. 2015).

- **Cognitive behavioral therapy (CBT)**, a psychosocial intervention that aims to reduce symptoms of mental health conditions, has been found effective in reducing anxiety and depression in several studies (Hoffman et al. 2012). Evidence also supports the link between improved mental health and improved employment outcomes (Banerjee et al. 2017; Conrad et al. 1998; Morgenstern et al. 2009). The MOMS Partnership, which provides group CBT to mothers with low incomes who are experiencing depressive symptoms, has been found to have positive impacts on participants' depressive symptoms and employment in non-experimental studies (Smith 2021a, 2021b).

Selecting specific programs

We consulted with federal and state policymakers, practitioners who administer or operate employment programs, researchers, intermediaries, advocates, and program developers to identify suitable programs for the NextGen Project. We asked them to consider the programs they knew of that had the features described in the previous section and were likely to meet our other selection criteria.

Once we identified potential programs, we spoke to program leaders. These discussions occurred mostly by telephone but sometimes in person. During the discussions, we assessed the program's interest in participating in the evaluation while also making our first assessment of whether a random assignment evaluation was feasible.

In these preliminary discussions with program leaders, we found that several programs were not suitable for the evaluation. Some programs we identified, such as many of the social enterprises we considered, were too small to support an impact evaluation. One social enterprise was large enough for the evaluation, but program leaders did not want to deny services to anyone eligible and hence could not participate in an RCT. We did not identify any sector-based or career pathways programs suitable for evaluation under the NextGen Project. Those we identified as potential candidates were either too small or had eligibility criteria that excluded people facing complex challenges to employment. Some large companies, such as CVS, operate apprenticeship programs for people with disabilities, but we did not identify a feasible way to conduct an RCT of these programs.

As a result of these preliminary discussions, we initially selected six programs for more extensive evaluability assessments: (1) Bridges, (2) Families Achieving Success Today (FAST), (3) IPS-AJI, (4) Philly WINs, (5) Western Mass MOMS, and (6) Progressive Employment. We are evaluating all but FAST and Progressive Employment.

Assessing evaluability

At each program, we conducted an evaluability assessment to (1) understand the program's theory of change, (2) determine whether the program was well implemented, (3) gain a deeper understanding of the feasibility of conducting a rigorous evaluation, and (4) determine the value of evaluating the program. We collected this information through discussions with leaders and staff in programs that used an approach aligned with OPRE's priorities and served (or could serve) the population of interest for the project. The study

team, OPRE, and SSA decided whether to pursue the program for evaluation based on the information gathered as part of the assessment process.

We did not proceed with Progressive Employment because we were unable to find a location at which an RCT was feasible. We conducted an evaluability assessment for the Progressive Employment model used by the Nebraska Vocational Rehabilitation (VR) agency. Historically, this agency could offer Progressive Employment only to those deemed to have the most severe disabilities. We discussed with the agency the possibility of testing the model with a broader population of VR participants. However, it subsequently received more funding and could no longer deny Progressive Employment to other VR participants. Hence, creating a comparison group was not feasible.

We provide details of the rationale for including the four programs in the NextGen Project—Bridges, IPS-AJI, Philly WINS, and Western Mass MOMS—in Appendices A–D. Although, based on the evaluability assessment, we began an RCT of FAST, enrollment in the study was lower than expected and the implementing organization had competing priorities. Thus, the implementing organization decided to discontinue participation in the evaluation.

Activities to prepare programs for evaluation

Once we had selected programs for evaluation, we conducted activities to prepare them for it. This process involved identifying the program's core components; providing technical assistance on implementing the evaluation and, in some cases, program services; and conducting formative evaluations.

Identifying core components

Core components are the essential functions and principles that define a program and are judged as necessary to produce outcomes in a typical service setting (Blase and Fixsen 2013). Documenting a program's core components for the evaluation was important for doing the following:

- **Understanding the logic underlying the program**. Understanding the core components helped us understand the hypothesized links between the activities and services offered and the outcomes expected.
- **Tailoring the participant data collection instruments, as appropriate**. We referred to the list of core components when tailoring the follow-up surveys to capture information about participation in the core activities and services offered, as well as the outcomes hypothesized to result from people participating in them.
- **Assessing implementation quality**. Knowing the core components of the program in detail gave us a standard against which to assess implementation quality, which we did as part of the descriptive study for each evaluation (see Chapter 3).
- **Building relationships with the programs.** The collaborative approach to documenting core components supported early relationship- and trust-building between the NextGen team and program staff. It also provided a benefit to programs that did not have documented core components.

The IPS model already had documented core components. For the other three programs, we identified the core components. To do this, we conducted a series of meetings with program leaders and staff to construct logic models. For Western Mass MOMS, we also collaborated with Yale University, the developer of MOMS Partnership. During these meetings, we examined participants' needs, how they are served, and their expected outcomes in the short and long term. We asked program leaders and staff to consider needed implementation supports, such as training, performance feedback, and resources, and organization culture. We also asked them to reflect on which policies and social and economic conditions could support or hinder effective program implementation.

With the program documented in detail, we then guided program leaders and staff through an activity to identify those elements of the program they deemed critical to achieving participant short- and long-term outcomes—the core components. The program leaders and staff largely made these decisions, but we supplemented the conversation with findings from the literature when possible.

Providing technical assistance

Technical assistance in preparing programs for evaluation included addressing issues related to the following:

- **Recruiting sufficient numbers of participants for the study**: This assistance involved learning how to identify the potential program participants—including screening criteria in some cases; how to inform potential participants about the study; next steps for those interested (for example, a warm hand-off to study staff); how potential participants would get to the physical (or virtual) location where services would be held; whether existing referral sources would be sufficient to achieve target sample sizes; and whether and how to recruit additional referral sources.
- **Defining the program and comparison groups**: We discussed with program staff where to insert the point of random assignment, what services would be available to the comparison group, and how to reduce potential for contamination of the comparison group.
- **Study procedures**: Technical assistance for these procedures covered which program staff would introduce the study to potential participants, what they would say about the study, how to field questions about the study from potential participants, how to administer consent and the baseline survey in RAPTER, how to conduct random assignment, how to inform participants about their study group assignment, next steps after random assignment for both study groups, how to document in RAPTER those services received by the program group, and routines and processes to maintain study group assignments.
- **Gaining buy-in for evaluation from program staff and the community**: We offered assistance about how to explain that the programs would be able to serve more participants because of the evaluation, highlighted the benefits of random assignment, identified key community groups to inform about the study, talked about the study with the broader community, and investigated what resources could support that communication.

In addition to the study team's technical assistance about the evaluation, the IPS Employment Center provides training and technical assistance in implementing the IPS model to the IPS-AJI programs. Yale University provides technical assistance about implementing Western Mass MOMS.

Formative evaluation

In addition to the technical assistance we provided to all the programs selected for evaluation, we conducted formative evaluations of two programs: Bridges and Western Mass MOMS. The formative evaluations included rapid learning cycles. Each cycle involved collecting data on a program component, analyzing them, and working with the program staff to refine implementation of the component. The cycles were sometimes repeated so data were collected on the refined implementation of the component.

For Bridges, the formative evaluation centered on quality and consistency of data program staff recorded about the services Bridges participants received. For Western Mass MOMS, formative evaluation focused on adding a new program component. See the appendices for further details.

3. Design of descriptive studies

The NextGen team is documenting and analyzing implementation of each of the four programs being evaluated as part of the NextGen Project through a descriptive study. This chapter describes the descriptive study design, which was pre-registered on the Center for Open Science's OSF registry. It begins with a discussion of the study objectives and research questions, and then describes our overall approach. Next, it describes the descriptive study data sources, concluding with a discussion of the analytic approach.

Descriptive study objectives and research questions

The two ultimate objectives of the descriptive study are: (1) to provide information useful to other organizations that might consider implementing a similar program, and (2) to interpret the impact findings. In addition, conducting a thorough descriptive study demonstrates to program leaders and staff that the evaluation team thoroughly understands their model (and is therefore well-positioned to interpret findings from the impact analysis). In some cases, findings from a descriptive study can also directly benefit the program being studied by providing information about the clients reached and the implementation strengths and challenges of the program.

To meet these objectives, for each program, the descriptive study:

1. **Describes the context in which the program was implemented**. Both the success of implementing a program and its effectiveness are affected by the context in which the program is implemented. Understanding the context is important in making sense of the findings. For example, the implementation and effectiveness of a program may differ if implemented by a different organization at a different time and in a different community. Context includes the characteristics of the population that is the focus of the program; the characteristics of the organization that is implementing the program; the partners involved in providing the services; as well as characteristics of the local community at the time of the evaluation—its labor market; the state of the COVID-19 pandemic; available transportation; and the other services available in the community to members of both the program and comparison groups.
2. **Documents the design of the program**. For each program being evaluated, we document in detail how it is intended to be implemented. If the impact study finds a program effective, other organizations might be interested in implementing it. The descriptive study report provides the information other organizations might want to consider in deciding whether their organization could feasibly implement a similar program. The report includes the core components and any principles that guide the program's implementation; other aspects of the program that may be important but were not deemed "core"; the program's logic model; and the criteria for program eligibility.
3. **Describes the program's implementation**. The descriptive study assesses whether the programs were being implemented as designed (that is, implemented with fidelity) at the time the descriptive study was conducted. Assessing fidelity to the program models helps inform future implementation, replication, and interpretation of impacts. Documenting both intentional and unintentional deviations from the design and why those deviations occurred will help future program implementers determine whether

changes are needed to the programs or how they are implemented. If we find differences between design and implementation, we will consider how those differences might be related to expected and actual program impacts. The fidelity assessment includes a description of how the implementers adapted a program design to fit the local context and environment; this is especially important for evaluations where an existing program model is being implemented in a new context (for example, implementing IPS for adults with justice involvement and implementing the MOMS Partnership model in western Massachusetts for the first time).

4. **Discusses potential implications for the impact study.** The descriptive study includes discussion of factors that may contribute to impacts being observed through the impact study as well as factors that may inhibit finding impacts. For instance, programs being implemented with high fidelity and participants receiving the intended dosage tend to support impacts; programs being too similar to other services in the community or low service take-up tend to dampen impacts.

The descriptive study of each program answers the research questions outlined in Exhibit 2. The research questions and the study objectives guided our development of data collection instruments, the data collection approach, and our data analysis. We asked program staff whether there were research questions that they wanted the descriptive study to address. While their suggested questions all fell under one of our existing research questions, we ensured that we collected data that could speak to their specific questions, so that the study findings are useful to them.

Exhibit 2. Descriptive study research questions

Describe the context in which the program was implemented
What are the needs and strengths of the population of interest?
What is the organization setting and location?
Who are the partners in providing the services?
What is the state of the labor market, COVID-19 pandemic, and other community factors that may contribute to the context in which the program is implemented?
What other services are available in the community to the program and comparison group members to assist them in becoming economically secure?
Document the design of the program
What are the program eligibility requirements?
What are the program's core components and guiding principles?
What are the other key elements of the program?
What is the program's theory of change?
Describe the implementation of the program
What are the characteristics of the program participants recruited and served?
What amount and type of services do program participants receive?
What strategies are used to implement the program?
How does the program's implementation differ from its design?
What factors help or hinder the implementation of the program?
What strategies appear to be needed to implement the program well?

Descriptive study approach

The descriptive study of each program follows five main principles to meet the study objectives and answer the research questions. Each descriptive study (1) is guided by a conceptual framework; (2) uses a focused approach to data collection grounded in the program's core components; (3) relies on data from multiple sources, collected over time; (4) incorporates perspectives of program staff and participants in data collection and analysis; and (5) follows a structured approach to assess fidelity. These principles are described below.

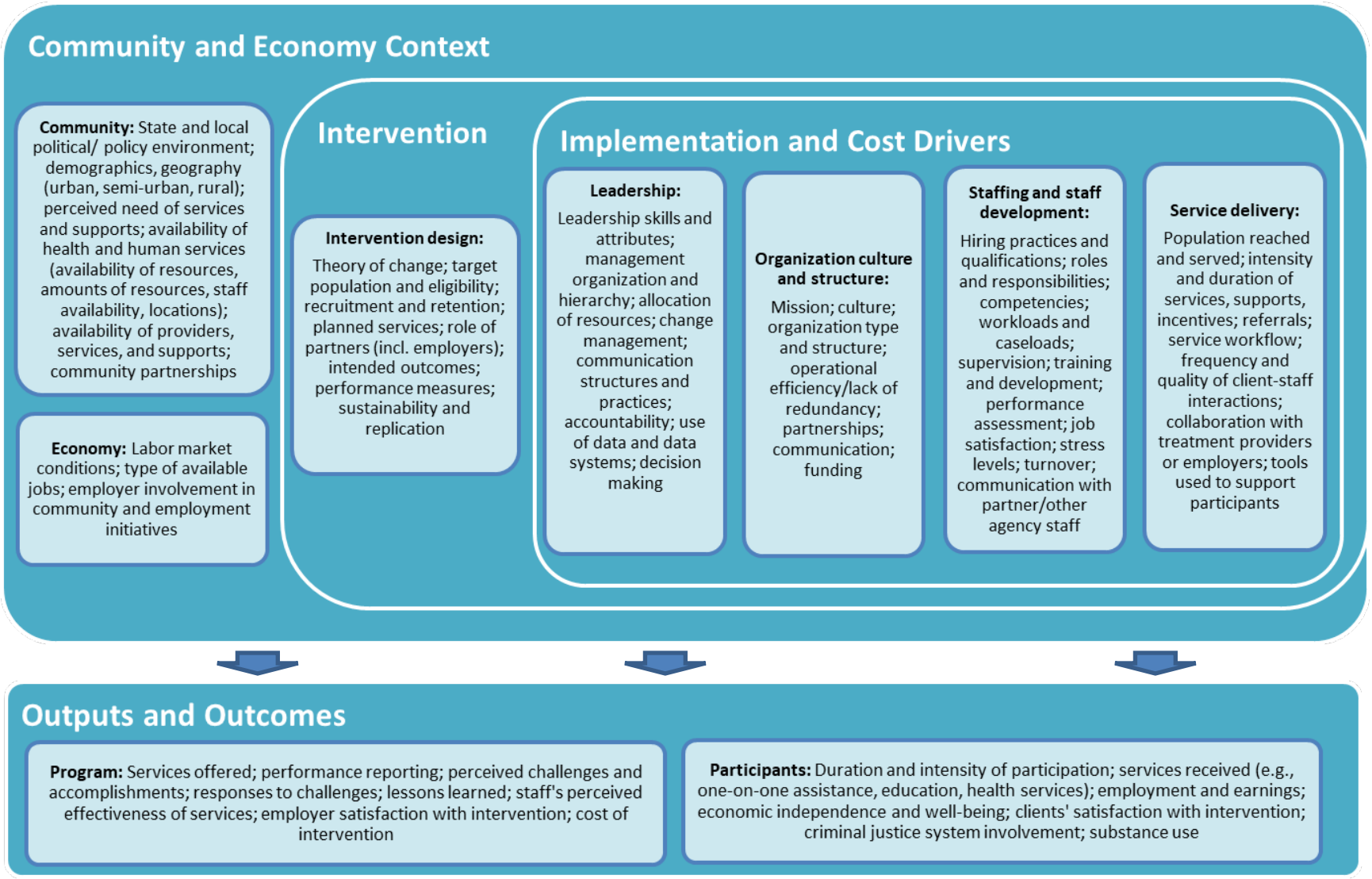
Guided by a conceptual framework

We developed a conceptual framework (Exhibit 3) to guide our descriptive study data collection and analysis. A conceptual framework shows the theorized relationship between different elements that we expect to influence program implementation and program and participant outcomes. We used the framework to develop our descriptive study data collection instruments and inform our analysis. Our coding scheme for the qualitative data incorporated the framework elements. Developing a framework before collecting and analyzing the data and then using it to guide our study helped ensure we asked the right research questions, collected the right data to answer them, and analyzed the data comprehensively. The literature on ecological models (Brofenbrenner 1995), which suggests there are multiple, inter-related layers of factors that influence a program, and existing implementation science frameworks, including the Consolidated Framework for Implementation Research (CFIR) (Damschroder et al. 2009), influenced our framework.

The conceptual framework theorizes that the implementation of the program as designed (the middle ring) is influenced by the (1) implementing organization's leadership, (2) its culture and structure, (3) available staffing and staff development, and (4) elements of service delivery (the innermost ring). Both the program as designed and its implementation also are influenced by the community and the economy in which it was designed and operates (the outer ring). Together, the community and economic context, the program, and how it is implemented determine the outputs and outcomes for both the program and its participants (box at the bottom).

Based on the framework, the study reports on the broad community and economic context; aspects of organizational leadership, culture, and structure; staffing and staff development; and the service delivery itself to understand how the programs were designed, how they were being implemented, and what factors support or hinder successful implementation.

Exhibit 3. Descriptive study conceptual framework



Use a focused approach to data collection grounded in the program's core components

As noted previously, we identified and documented the core components of each program while preparing for the evaluation. Upon the start of study enrollment, we began routinely monitoring and collecting information about implementation of the core components during evaluation technical assistance conversations; through data collected via RAPTER; and during program leader, staff, and participant interviews for the descriptive study. Documenting the core components as designed early in the evaluation and monitoring implementation throughout the evaluation period helped us correctly prioritize data collection for the descriptive study.

Rely on diverse data sources collected at different points in time

To obtain a full picture of the design and implementation of each program, we collected data from program staff and leaders; partner staff (including organizations that refer potentially eligible applicants and partners involved in the design of the programs); study participants; and, for some programs, employers. Our methods for collecting data included the following:

- Ongoing conversations with program leaders and staff
- Semi-structured and in-depth interviews with staff and participants
- Surveys of leaders, staff, and participants
- Observations of program activities
- Requests for program staff to enter information on program group members' service receipt in RAPTER or the program's management information system
- Review of program documents and other relevant literature

Formative evaluation activities and third-party fidelity reviews were also data sources for some programs. Each data source is described in more detail below.

Data collection began when each program was first selected for the project and will continue throughout service provision. Once the program reached a steady state of implementation, and at least 75 program group participants had been enrolled for three months, we conducted our descriptive study interviews and wrote a report representing a snapshot of implementation. We will account for any information collected after this point in subsequent reports. This approach helps ensure that our analysis is inclusive of all of the information we gathered from the programs throughout the start-up period and as the programs evolved toward steady state. Collecting information consistently during our early work with each program also helped the data collectors assessing implementation dig more deeply into the how and why of implementation during interviews and observations.

Incorporate perspectives of program staff and participants in data collection and analysis

The descriptive study includes the perspectives of program staff who were providing program services and participants who were receiving them. We consulted with program

staff about issues or nuances to be aware of when conducting interviews with staff and participants. We also asked program staff for their suggestions about which types of staff and which employers and partner agencies we should interview. In addition, we asked program staff to review the in-depth interview discussion guide to check that the questions and language used were culturally appropriate and relevant for the population they served. After conducting the first in-depth participant interview at each program, we gathered informal feedback from the participant about whether any of the questions or language used was confusing and probed about any questions answered in a way that was unexpected to the interviewer.

We presented our initial findings to program staff before drafting the descriptive study reports. We explored whether the findings were aligned with staff expectations and asked for help with interpreting findings, as needed.

Follow a structured approach to assess fidelity

A key element of the descriptive study is understanding whether the programs were being implemented with fidelity to their designs at the time of the descriptive study data collection. IPS has preexisting specific, validated fidelity measures and assessment processes; the IPS Employment Center conducted fidelity reviews for the IPS-AJI program. For programs without preexisting fidelity measures and assessment processes, we examined implementation of their core components as designed.

Descriptive study data sources and uses

To answer the descriptive study research questions, we relied on multiple data sources that incorporated diverse perspectives, described in the previous section. Exhibit 4 maps the research questions to each of the data collection sources. The rest of this section describes our approach for tailoring data collection instruments to each program and then details the information each data source provided to address these research questions.

Exhibit 4. Descriptive study research questions and their data sources

Research question	Data collection sources										
	Ongoing conversations with program leaders and staff	Formative evaluation	Interviews with program leaders and staff	Interviews with partners and employers	Staff and leadership surveys	Observations and job shadowing	In-depth interviews with participants	Service receipt data	Program documents, literature, and data	Baseline and follow-up surveys	Third-party fidelity reviews
Describe the context in which the program was implemented											
What are the needs and strengths of the population of interest?	✓	✓	✓	✓		✓	✓			✓	
What is the organization setting and location?	✓		✓								
Who are the partners in providing the services?	✓		✓	✓							

Research question	Data collection sources										
	Ongoing conversations with program leaders and staff	Formative evaluation	Interviews with program leaders and staff	Interviews with partners and employers	Staff and leadership surveys	Observations and job shadowing	In-depth interviews with participants	Service receipt data	Program documents, literature, and data	Baseline and follow-up surveys	Third-party fidelity reviews
What is the state of the labor market, COVID-19 pandemic, and other community factors that may contribute to the context in which the program is implemented?	✓		✓	✓					✓		
What other services are available in the community to the program and comparison group members to assist them in becoming economically secure?	✓		✓	✓			✓		✓	✓	
Document the design of the program											
What are the program eligibility requirements?	✓		✓								
What are the program's core components and guiding principles?	✓		✓						✓		
What are the other key elements of the program?	✓		✓	✓	✓	✓	✓				
What is the program's theory of change?	✓		✓								✓
Describe the implementation of the program											
What are the characteristics of the program participants recruited and served?	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
What amount and type of services do program participants receive?			✓	✓		✓		✓	✓		
What strategies are used to implement the program?	✓	✓	✓		✓	✓	✓	✓	✓		✓
How does the program implementation differ from its design?	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
What factors help or hinder the implementation of the program?	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
What strategies appear to be needed to implement the program well?	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Tailoring instruments to each program

We tailored the interview protocols and observation data collection instruments to reflect what the NextGen team already knew about the programs while also probing about topics and challenges that had already come to light. Pursuing these additional topics and exploring challenges helped us dig more deeply into implementation issues of interest for each program and prevented interviewers from asking staff about descriptive information

we already collected. Instead, during interviews, we focused on understanding how staff and partners were implementing the programs to assess fidelity to the core components and other elements of program design, and to understand the implementation context, nuances, challenges, and lessons learned. We used service receipt data, responses to staff and leadership surveys, results from formative evaluations (described below) and fidelity reviews, and notes from ongoing conversations to identify areas that required additional exploration during staff and leadership interviews. Similarly, we tailored the in-depth interview discussion guide to ensure that interviewers probed about the core components of each program and any contextual factors on which participants may have perspectives.

Data collection sources

Ongoing conversations with program leaders and staff

We collected information about a program's design, core components, and implementation during upfront and ongoing evaluation planning and continue to do so during technical assistance and monitoring conversations. To ensure the information is documented, and for ease of analysis, we input key information from the conversations into a database on an ongoing basis. This process allows the data to be easily retrieved during data analysis.

Formative evaluation

As part of the evaluation planning for Bridges and Western Mass MOMS, we conducted formative evaluation activities to help prepare the programs for the impact evaluation. These activities included improving data collection approaches (Bridges) and pilot testing aspects of the program (Western Mass MOMS) to refine implementation. We included the results of these evaluations as a data source for the descriptive study to help describe how the programs were implemented at the time of descriptive study data collection. Details about the formative evaluations are included in the appendices of the relevant programs.

Semi-structured interviews with program managers and staff

We interviewed program managers and staff in person or by telephone using semi-structured interview protocols. These discussions explored elements of the program's design, staffing, service provision, partnerships, and other details necessary to understand them and their context. Exhibit 5 presents the main topics discussed during these interviews.

We selected program staff and leaders purposively for discussions using organizational charts and information on each employee's role at the organization. Types of program staff included the lead administrator responsible for the program; supervisors; frontline case managers or employment specialists; job developers; mental health clinicians; teachers or group facilitators; and any other staff responsible for providing employment services.

We determined the number and type of leaders and staff interviewed, and whether interviews would be conducted in person or by telephone, separately for each program, depending on the program design, number of program locations, and number of staff. To the extent possible, we aimed to include leaders and frontline staff from all program locations participating in the project. If many staff filled the same role, we selected a sample of the staff to interview. We aimed to create a diverse sample, considering their time in the position and demographic characteristics, such as age, race, and gender. Interviews lasted between 45

and 90 minutes, and were individual or small-group discussions (no more than three staff people with similar roles). Details of each program's descriptive study design are included in the appendices.

Exhibit 5. Program leader and staff discussion topics

Topic	Example subtopics
Respondent background	Job title; tenure with organization
Economic context	Local economic conditions; economic challenges
Structure and staffing of the program	Mission of implementing organization; number and qualifications of staff
Program development and target population	Characteristics of target population; motivation for developing the program
Services provided by the lead organization	Recruitment and outreach; participant flow through services
Services provided by partners	Characteristics of partner agencies; communication with partners
Engagement with employers	Roles of employers; how employers are recruited
Program participants and counterfactual	Program participants' backgrounds, strengths, and challenges; other available programs and supports for the target population
Program monitoring and oversight	Performance targets and outcomes; extent to which program has met goals
Sustainability and lessons learned	Program costs and funding; key successes and challenges

Semi-structured interviews with program partners and employers

We interviewed staff of partners who refer applicants or were involved in the program design, as well as employers, as applicable, in person or by telephone, using semi-structured interview protocols. Exhibit 6 presents the main topics discussed during partner interviews; Exhibit 7 presents the main topics discussed during interviews with employers. We selected the partners and employers, and the specific staff with whom to speak, based on their level of involvement with the program and its participants, and their understanding of the community in which the program operates.

Exhibit 6. Partner staff discussion topics

Topic	Example subtopics
Respondent background	Job title; tenure with organization
Economic context	Local economic conditions; economic challenges
Structure and staffing of the program	Mission of implementing organization; number and qualifications of staff
Program development and target population	Characteristics of target population; motivation for developing the program
Services provided by the lead organization	Recruitment and outreach; participant flow through services
Services provided by partners	Characteristics of partner agencies; communication with partners

Topic	Example subtopics
Program participants and counterfactual	Program participants' backgrounds, strengths, and challenges; other available programs and supports for the target population
Program monitoring and oversight	Performance targets and outcomes; extent to which program has met goals
Sustainability and lessons learned	Program costs and funding; key successes and challenges

Exhibit 7. Employer staff discussion topics

Topic	Example subtopics
Respondent background	Job title; tenure with organization
Employer characteristics	Mission; for-profit or nonprofit status; number of employees
Experiences with program	How employer got involved; role employer plays as part of the program
Employees hired through program	Hiring process; pay; supervision
Satisfaction and lessons learned	Benefits derived from working with the program; satisfaction working with program

Staff and leadership surveys

We asked all staff and leaders at the programs to complete a web-based staff or leadership survey collecting information about their professional backgrounds, job responsibilities, and perceptions of the program. Surveyed staff included program managers; supervisors; frontline case managers or employment specialists; job developers; mental health clinicians; teachers or group facilitators; and any other staff responsible for providing employment services. Exhibit 8 presents a summary of the topics included in the surveys.

All of the survey questions were closed ended. This means that the information collected was more systematic and standardized than that collected during interviews or via conversations. The survey also enabled us to collect information on topics that staff may be uncomfortable discussing, such as their perceptions of the quality of the program or organizational practices. Surveying all staff and leaders is appropriate to gain a broader perspective on these topics than we could elicit through semi-structured interviews.

Exhibit 8. Staff and leadership survey topics

Topic	Example subtopics
Respondent background	Job title; years of experience
Staff responsibilities and contact with participants (staff only)	Program responsibilities; hours in a week spent on various tasks
Responsibilities and decision making (leadership only)	Hours in a week spent on various tasks; training received
Perceptions of program	Challenges participants face; helpfulness of services
Program organizational practices	Extent to which staff make an effort to get to know participants; extent to which staff have the required skills

Observations of program services and job shadowing

We observed select program services and activities, such as orientation sessions, classes, and case management meetings, either in person or virtually. When feasible and appropriate, we also conducted job shadowing, in which study team members shadowed program staff members as they conducted program activities. The observations and job shadowing helped us better understand program implementation and whether activities were implemented as designed.

In-depth interviews with program participants

We conducted in-person, one-on-one interviews with select study participants to gather in-depth information about the participant experience. In-depth interviews with participants allow the participants to voice their experiences and provide rich information about participants' lives, their work experiences, time participating in the programs, and opinions of the programs. These interviews provide "stories" that make the findings from the descriptive studies and, later, the impact studies more meaningful. The information also informs our understanding of whether the program was implemented as planned and suggests possible refinements. In addition, the in-depth interviews helped us understand the population served by the programs and the communities in which they lived in more detail—both important aspects of the program context.

We aimed to interview about 12 participants per program for up to two hours, either in person or virtually. We attempted to recruit program group members who were randomly assigned at least six months before the interviews. The interviews were conducted in English or Spanish depending on the preference of the study participant. All study participants who completed an interview were given $60.

We trained all interviewers regarding the interviews' goals and how to use techniques designed to obtain the most information. For example, we trained them on ways to evoke detailed narratives, using probes such as "Tell me the story about that," "What happened then?," and "Where do things stand now?" Interviewers learned the protocol so they could use a conversational tone, allowing the participant to lead the conversation while ensuring they covered all topics of interest in the protocol. Using a conversational tone and following the participant's lead allows interviewers to develop rapport with the participants and help them feel more comfortable sharing their perceptions and experiences. All interviewers were trained in unconscious bias and cross-cultural understanding before beginning data collection.

Major topic areas for the interviews are presented in Exhibit 9.

Exhibit 9. In-depth participant interview topics

Topic	Example subtopics
Respondent background	Names and ages of children; where participant grew up
Experience with work	Employment history; employment goals
Experiences with the program	Whether still participating
• Initial experience	First impressions of program; motivation to apply

Topic	Example subtopics
• Relationships with program staff	Frequency and content of communications with staff; opinions of relationship with staff
• Program participation	Activities completed; opinions of program activities
Reflections on program experience	What about program has been helpful; advice for potential participants

Service receipt data from RAPTER and/or program management information systems

Western Mass MOMS staff use RAPTER to record information about program group members' participation in the program. The other programs (Bridges, IPS-AJI, and Philly WINs) already collect data on service receipt through their own management information system. We used these data to describe the service receipt of program group members—and, in some instances comparison group members—including type of service received, duration, and mode. Exhibit 10 presents the types of data that are collected in RAPTER or programs' management information systems.

Exhibit 10. Service receipt data captured in RAPTER or program management information systems

Topic	Subtopics
Program enrollment	Enrollment date and referral source
Service contacts	Type of service received; date, mode (in-person versus virtual), and length (minutes) of service
Group events (such as classes, workshops)	Purpose; date; location; participant attendance
Collaboration with employers and partners	Partner; reason for meeting; date, mode, and length (minutes) of meeting
Financial and in-kind support	Type; value; reason for providing; date
Referrals	Date; purpose of referral; referral agency
Case status	Whether participant is active in program, completed, or dropped out (and reason for dropping out)

Program documents and data

For each program, we collected and reviewed policy and procedures manuals, staff training materials, recruitment materials, curricula, forms used to document program activities, and other relevant documents that contain key information about the program as designed and its operations. In addition, we reviewed available literature and data about the communities in which the programs operate, and the population served, to understand more fully a program's context and the culture in which it operates.

Baseline and follow-up surveys

The baseline and, to a limited extent, first follow-up surveys include information relevant to the goals of the descriptive study. The baseline survey, administered to program-eligible individuals just before random assignment, collects information about individual demographic characteristics, housing status, public benefit receipt, social support,

employment status, goal setting, criminal justice system involvement, disabilities, and physical and mental health (Exhibit 13 in Chapter 5). The first follow-up survey (but not the second) asks program group members about their opinions of the program. The NextGen team used the baseline survey data to describe the program group in the descriptive study. Data from the first follow-up survey about program group members' opinions was limited because relatively few participants had been surveyed at the time of the other descriptive study data collection.

Third-party fidelity reviews

As part of program implementation, the IPS Employment Center conducts fidelity reviews of the IPS aspects of IPS-AJI using validated fidelity measures. The fidelity measures and the process for these reviews were established before this project and have been found to be psychometrically valid in several studies (Bond et al. 2011). At least one fidelity review was completed for each mental health center in the evaluation of IPS-AJI. The results of the fidelity reviews were an additional data source for the descriptive study. Additional details about these reviews are included in the program's section of the appendices.

Descriptive study analytic approach

We analyzed the qualitative and quantitative descriptive study data using descriptive, comparative, and thematic analysis approaches. The first step in our analysis was to describe the programs as designed. This includes:

- Program origins and background
- Logic model and core components
- Program services
- Services available to the comparison group
- Target population
- Implementing organization
- Providers, partners, and other stakeholders
- Funding sources
- Data, performance, and accountability
- Community and economic context

To analyze the qualitative data for each program, we used a thematic analysis approach (Braun and Clarke 2012; Deterding and Waters 2021). This involved the following steps:

1. **Determining initial codes.** We developed initial analytical codes for potential answers to each research question. These analytical codes were based on internal discussions among team members and knowledge of the literature on implementation of employment programs.
2. **Coding the data.** We tagged portions of the data with analytic codes that highlighted important aspects of their content or interpretations of the content. We sometimes

needed to add or change some of the initial codes as we reviewed the data. The data relevant to each code were collated.

3. **Searching for themes.** We analyzed the collated data to determine the themes that are reflected in the data. We identified where the codes were similar or connected, and where they suggested a trend or pattern. This involved assessing similarities or apparent conflicts in the responses and determining if different types of respondents shared differing perspectives or experiences.
4. **Reviewing potential themes.** We cross-checked the draft list of themes with the initial codes and, as necessary, the original recordings or interview notes to make sure that the themes correctly reflected what was described by the respondents. The study team made sure that themes emerging across different research questions were not contradictory. In cases where contradictions were found, the team further explored the data to address the conflicting themes.
5. **Defining and naming findings.** For each theme, we documented a narrative story of a finding that captures how the theme is explained or defined as well as how the data provide the evidence to support that finding. This included respondent quotes that exemplify the theme or researcher interpretations of the data.

The analysis of quantitative data supplements the discussion of themes in the qualitative data. For the participant service receipt data, we calculated descriptive statistics, including frequencies, percentages, and averages, to describe trends in services received, average dosage, and trends in dosage over time. We also calculated descriptive statistics of the baseline survey data to describe characteristics of the program group members when they enrolled in the study. Similarly, we computed descriptive statistics, such as means and frequencies, of the staff and leadership survey data to help describe characteristics of the staff and leaders, the organizations in which they work, and staff perceptions of the programs.

We analyzed the coded qualitative data and descriptive statistics to compare how the program is being implemented with how it was designed to be implemented. We reviewed the data to identify themes related to factors that seemed to be impeding or facilitating implementation and any challenges the programs faced, solutions used, and lessons learned.

To help ensure the quality of the descriptive study analysis across data sources, we held internal team meetings to debrief about the emerging themes and findings from the data to make sure there was agreement, and that emerging themes across different research questions were not contradictory. Thus, to the extent possible, the findings draw from the analysis of multiple data sources, including staff and participant interview data, observational data, and program documents. In addition, more than one study team member analyzed the data; this approach brought in different perspectives and helped confirm emerging themes and findings (Deterding and Waters 2021). Finally, a quality assurance reviewer not involved in the analysis reviewed the findings and conclusions to ensure they were well documented and supported by the data.

4. Design of cost studies

For each of the four programs being evaluated as part of the NextGen Project, we provide an estimate of the cost of providing the services. This chapter describes the cost study design. It begins with a discussion of the study objectives and research questions, and then describes our overall approach. Next, it describes the cost study data sources, concluding with a discussion of the analytic approach.

Cost study objectives and research questions

The objective of the cost study is to estimate the cost of the program—both overall and per participant. This is important because decisionmakers must have an estimate of a program's costs in order to make informed resource allocation decisions. For example, providers and policymakers not currently implementing the program need to know how much it might cost before deciding whether to try to replicate it in their setting. Cost information also provides context for interpreting the magnitude of the estimated impacts. For example, if earnings impacts are $200 per month for 6 months, decisionmakers might draw different conclusions about the size of this impact if the cost of the intervention were $500 per participant versus $3,000 per participant.

The research questions for the cost study are:

1. What is the total annual cost to implement the program?
2. What are the components of the cost? For example, how much of the cost is for labor, other direct costs (such as for staff travel or bus passes), payments to participants, and overhead (such as facilities and management)? How many volunteers and other in-kind services are involved in program implementation?
3. How does the cost break down by funding source?
4. What is the total cost per participant per month?
5. What is the total cost per participant?

Design decisions

In conducting a cost analysis, some design decisions need to be made. This section describes our design decisions and their rationale. As much as possible, the decisions are guided by making the cost analysis useful to potential program managers, funders, and policymakers.

Perspectives

The costs of a program can be borne by three main categories of people or organizations:

1. Government agencies that provide funding or other resources (such as volunteers or space) for services. This cost is ultimately borne by the taxpayers.
2. Other nongovernmental organizations and people that provide funding or other resources for services.

3. The program participants themselves. Participants may bear the cost of tuition or fees, for example. They may also receive money as incentives for attending a session, which are negative costs, or benefits.

Policymakers, prospective programs and funders, and participants will all take different perspectives. A policymaker considering whether a program is cost-effective considers all the costs (the sum of all three categories above). This is sometimes referred to as the perspective of society as a whole. An organization considering whether to implement the program will need information about the amount of the costs borne by government agencies and other nongovernmental organizations and people (the sum of the first two categories above) to know how much they need to fund the program. A participant will only consider the cost to the participant.

Estimating the total cost from the cost of each component

We use the "ingredients" approach to estimate the total cost of each program (National Academies of Sciences, Engineering, and Medicine 2016; Levin and Belfield 2015). Under this approach, we (1) list each resource required to deliver a program, (2) determine the monetary value of each, and (3) sum these values to estimate total annual costs. We value all resources—labor, facilities, equipment, and overhead—required to deliver it, including those that may not appear on an organization's expenditure records (for example, volunteers and other in-kind resources).

Annual cost, cost per participant-month, and cost per participant

Program managers often think about the cost of a program in annual terms—how much do they need to allocate to fund the program for a year. They may also be interested in how much it costs per participant per year or per month. This helps them understand how the costs may change if they change the number of people served. The benefits of a program—such as the increase in earnings—are measured per participant. Hence, a policymaker comparing the costs of the program to its benefits needs to know the cost per participant.

Services to include in the cost estimate

For some programs, the boundary around which services are part of the program and which are additional services is blurred. Our approach is guided by the services available to the comparison group. For example, for the evaluation of IPS-AJI, program group members are offered IPS employment services integrated with mental health services whereas the comparison group members are offered only mental health services. Because both groups can receive mental health services, we will compute the cost of only the IPS-AJI employment services. This is the most relevant cost from the perspective of the policymaker, who will be comparing the difference in benefits received by the program group members and comparison group members (such as the additional earnings). It is also the most relevant for many providers who will already be providing mental health services and are considering adding IPS employment services.

Not including the costs of the evaluation

Programs incur costs while participating in an evaluation and these costs are not relevant for program managers, participants, or policymakers considering the costs of the program without an evaluation. Hence, we estimate the costs of the evaluation and subtract them from the total costs. Evaluation costs include the cost of administering consent and the baseline survey to study participants; entering additional service receipt data for the study into RAPTER or a program management information system; and program staff meeting with the study team throughout preparation and execution of the study. We also account for the cost of recruiting twice as many people as necessary for the program in order to fill the comparison group. Generally, we do this by subtracting out half the salary of staff who recruited participants for the study.

Including the costs of required program technical assistance

Some programs receive technical assistance on implementing their program. To the extent this was provided as part of the evaluation, we do not include those costs. However, we include in the costs of the program any costs associated with receiving ongoing technical assistance that is critical for successfully implementing the model, according to the program developers. This includes, for instance, technical assistance from the IPS Employment Center to support IPS-AJI and from Yale University to Western Mass MOMS.

Volunteers, donations, and other in-kind contributions

Some programs use staff and other resources that they do not pay for that we will include to get a complete picture of what it would take to replicate or sustain the programs. For example, staff from the Massachusetts Department of Transitional Assistance, not paid for by Western Mass MOMS, provide some services as part of the program. Other programs may receive donations of equipment or items to provide program participants. We will estimate the dollar value of these contributions.

Period of time over which to measure costs

Five considerations guide our selection of the appropriate period of time over which to collect cost information in each program:

1. **Represent steady state as nearly as possible.** The most relevant costs for program managers and policymakers are the costs over a period of time during which services were offered in the way they are intended to for the foreseeable future and the programs operated at or near capacity. This would suggest avoiding a time period in which programs were newly implemented and not operating at capacity. It also suggests avoiding the period during the COVID-19 pandemic when there were serious service disruptions and adaptations. That said, programs are constantly changing and adapting to meet changing needs of their participants. Thus, steady state is a relative term.
2. **Include the period during which study participants received services**. To reflect the economic and other conditions during the study period as nearly as possible, the cost data should include the period during which study participants received services.

3. **Include a calendar year**. Seasonal differences in the number of participants are common, which leads to seasonal differences in the cost per participant. Ideally, the reporting period would cover a calendar year to reflect those seasonal fluctuations.
4. **Align with the program's fiscal year**. It is often easier for a program to provide cost data when it aligns with its own fiscal year.
5. **Align with the timing of the descriptive study analysis**. Program managers would benefit from receiving information about program costs at the same time as we report on the descriptive study so it can inform their decision making.

The appendix for each program describes how these considerations play out in each program and the proposed timing of the cost study.

Cost study data collection

The program managers are the primary source of data for the cost study. We collected data from them using a Microsoft Excel workbook. The workbook contains tabs for different types of costs: total costs; costs of staff and volunteers; purchased services (for example, services contracted by the program from a vendor); in-kind donations and volunteers; and overhead costs (for example, management, computers and telephones, and facilities). The workbook collects information about how many participants are served during each month of the reporting period and the program's funding sources.

While some programs can isolate the cost of providing services to those services provided to the study participants, other programs can only provide costs for services provided to both study participants and other program participants. For example, Bridges reported data for the program as a whole, and not just those offices participating in the evaluation. For these programs, we collected data on the total number of participants served by the program, whether or not they participated in the study.

We worked with the programs to tailor the cost workbook so it best reflected the program's operations and the structure of their existing data. For example, when staff for a given program have responsibilities to other programs as well, we adjusted the workbook instructions so it was clear that only the time spent on the program being evaluated through the NextGen Project should be included.

The workbook contains definitions of each requested element, examples, and detailed instructions for completion. In addition, we held phone calls with the programs to explain the purposes of the data collection and specifics of the request. We monitored for quality and consistency by reviewing completed workbooks, checking for completeness and internal consistency.

We used information reported in RAPTER or programs' own management information systems to determine the number of study participants in the program group in each month in the reporting period and the average months of participation in the programs.

Cost study planned analytic approach

The approach involves estimating total annual costs and its components and the funding sources to cover it, total cost per participant per month, and total cost per participant.

Total annual costs. To calculate the total annual cost of delivering the program over the reporting period, we sum the value of all resources from all sources used. We make appropriate adjustments if needed, such as annuitizing the costs of equipment. We calculate the percentages of the total cost that come from each of the main cost categories: staff, volunteers and donated goods, purchased services, facilities, equipment, and other overhead. We list in-kind resources and also estimate a dollar value of those resources using information on the market cost of similar resources. We also report how the costs are covered by each funding source (that is, government and nongovernment funds).

Total cost per participant per month. To estimate the cost per participant month, we divide the total annual costs by 12 to obtain average monthly costs and then divide that by the average number of program participants per month. We asked the programs to report the number of total participants being served (not just the new participants) in each month during the reporting period.

Total cost per participant. To estimate the cost per participant, we multiply the cost per participant per month by the average number of months study participants received services from the program.

5. Design of impact studies

The objective of each impact study is to determine the effectiveness of each program being studied under the NextGen Project in helping people who face complex challenges to employment become economically independent.

This chapter describes the research questions to be addressed by the impact studies, the study design, how the studies are being implemented, the data to be used in addressing the research questions, and the methodological approach to the analysis. It ends with a short description of how the impacts will be compared with the costs of the program in a benefit-cost analysis. Evaluation-specific appendices include design details for each program that supplement the broad, cross-evaluation approach discussed in this chapter.

Impact study objectives and research questions

The impact studies address research questions that are aligned with the central goals of the NextGen Project and fine-tuned based on each program's logic model.

To be included in the NextGen Project, the programs were expected to have an impact on employment and promote economic independence—in particular, a reduction in the need for SSI. This reflects SSA's interest in better understanding the types of programs that effectively connect or reconnect potential SSI applicants to work before they apply for benefits. Hence, the key research questions for each program include those related to its impact on employment and economic independence. However, the programs being studied under the NextGen Project differ in the mechanisms through which they are expected to lead to improved employment and economic independence; therefore, other targeted outcomes, the measure of employment and economic independence, and when the impacts are expected to occur vary by program. For example, Bridges aims to affect employment in the short-term while Western Mass MOMS aims to affect mental health outcomes as well as employment and earnings.

The key research questions address the extent to which the programs being studied under the NextGen Project improve outcomes of interest; they include the following:

- **Does the program affect participants' employment outcomes?**

  We will examine outcomes such as earnings, employment, job retention, and job quality.

- **Does the program affect participants' economic independence?**

  For all programs, we will examine whether they reduce the need for SSI. For some programs, we will examine whether they reduce the need for TANF and Supplemental Nutrition Assistance Program (SNAP).

- **Does the program affect the amounts and types of services participants receive?**

  For all programs, we will examine whether they increase receipt of employment services. For some programs, we will examine whether they increase use of mental and physical health services.

- **Does the program affect participants' health and other outcomes?**

  For some programs, we will examine whether they improve participants' mental health conditions degree of social support, involvement with the criminal justice system, and other outcomes.

Addressing these research questions will inform our assessment of the effectiveness of the programs being studied under the NextGen Project. In addition to these key research questions, the impact studies will address the following additional research questions to shed light on how the programs work, and for whom:

- Are programs more effective for some groups of participants than others?
- To what extent do impacts on intermediate outcomes such as mental health and social support explain the long-term impacts on employment?
- Are programs effective for study participants who receive different amounts of program services?

Impact study experimental design

The experimental design involves participants who are eligible for the program services and have consented to participate in the study being randomly assigned to one of two groups: (1) a program group offered the program services or (2) a comparison group not offered those services but free to seek other services available in the community. With this design, the research groups should be similar in their characteristics before receiving the program services. Differences in observed outcomes thus can be attributed to the programs.

We tailored the evaluation design to each program. We worked with program staff to determine (1) when to insert random assignment within the program intake and service flow, (2) what services would be available to the comparison group members, and (3) ways to minimize potential threats to the validity of the research design.

Point of random assignment

The point in the intake process in which random assignment is conducted affects the strength of the evaluation and the research questions it addresses. We selected the point of random assignment together with program staff. In selecting the point of random assignment we considered: (1) study participants' interest in the program at random assignment—our aim was to reduce the proportion of program group members who do not end up participating in the program; (2) the investment in time, money, and emotion on the part of potential program participants before random assignment—our aim was to reduce harm to those assigned to the comparison group; and (3) comparison group members' receipt of program services prior to random assignment—our aim was to reduce the amount of program services received by the comparison group. The appendices discuss the selected point of random assignment for each program. As described below, in all programs random assignment occurs after participants have consented to participate in the evaluation, provided some identifying and contact information, and completed the baseline survey.

Comparison conditions

The programs participating in the NextGen Project refer comparison group members to other services in the community. Program group members also have access to these services. Employment services are available in the communities—for instance, from American Job Centers, public assistance programs such as TANF and SNAP, and community-based organizations. In the IPS-AJI evaluation, the comparison group is offered mental health services provided by the organization offering IPS-AJI. The appendices discuss the comparison conditions in each program's evaluation.

Preventing and monitoring crossover and contamination

Impact estimates can be biased in situations in which the comparison group is directly or indirectly affected by the intervention. One such situation, referred to as crossover, occurs if comparison group members receive program services that only program group members are supposed to receive. Another such situation, referred to as contamination, can occur when members of the comparison group are affected indirectly by the intervention. Contamination might occur if, for example, introducing new program services causes changes to the organizational culture generally around service provision. This may cause staff who serve the comparison group to change their approach to service provision in ways that they would not have in the absence of the new program, even though they are not delivering the new program services themselves. Finally, program staff may be tempted to assist comparison group members in finding other services in the community and so they may receive more services than they would in the absence of the study.

We took steps to preserve the validity of each evaluation's impact estimates from crossover and contamination. For instance, we designed the evaluation so that program staff do not have contact with comparison group members after they have been randomly assigned and referred to other services. We also designed the random assignment procedures so intake staff check that all new potential participants have not already been enrolled in the study, to preserve the random assignment groups. We adopted simple procedures, provided clear written instructions in the study procedure manual, and provided detailed training to program staff. If we identify through ongoing monitoring that compliance is an issue, we will provide additional training.

Addressing low program participation

Maximizing the extent to which the program group receives program services is another important step in preserving the policy relevance of the evaluation's impact estimates. If members of the program group do not take up program services, the difference between the program and comparison groups will be smaller, making it harder to detect impacts for a given sample size. If we were to find no impacts, we would not be able to determine whether this result is because program services are not effective or because program group members did not receive sufficient doses of the program to generate impacts.

We have attempted to minimize this problem by (1) selecting a point of random assignment in which the study participants have shown interest in the program by, for example, attending an in-person session; (2) selecting programs that demonstrate an ability to engage

participants; and (3) monitoring attendance and providing technical assistance to help programs boost attendance. We also plan to conduct dosage analysis to estimate how effective the intervention is for those who receive different amounts of program services, as discussed in greater detail in a later section.

Conducting and monitoring random assignment

Implementing random assignment involves program staff obtaining consent, administering the baseline survey, and conducting random assignment using RAPTER. Program-specific evaluation manuals document the process. This section discusses the logistics of how random assignment occurs.

Informed consent

After program staff have told the applicant about the program and the study and answered their questions, they read the consent statement. Bridges includes participants under the age of 18. In this program, informed consent is also collected from the participant's parent or guardian.

The consent form describes the study and the implications of the study for the participant (such as the administrative data that we will collect and the surveys we will ask them to complete), how to contact the NextGen team, and how to withdraw from the study. We tailored the consent forms for each program to reflect the services available to the comparison group and the plan to collect administrative data, among other aspects.

Participants who do not consent will not be enrolled in the study and cannot receive program group services for a period aligning with the timing of the second follow-up survey, which varies from 18 to 21 months, depending on the program.

Random assignment

Program staff conduct random assignment using RAPTER. RAPTER checks that the applicant has not already been randomly assigned, and if not, will assign the applicant to the program group or comparison group with equal probability.

After informing participants about their study group assignment, program staff provide them with a study packet designed to establish their engagement with the study. This packet includes a copy of the consent form, as well as information about the follow-up survey and, to help the participant remember the study, a small gift with the name of the study (such as a drawstring bag or a screen cleaner).

Monitoring random assignment

To ensure that the study design is faithfully implemented and the integrity of random assignment is maintained, we monitor random assignment. We review reports from RAPTER to show the percentage of people who have been assigned to each group. We regularly discuss with program staff how they are conducting random assignment and how they interact with comparison group members. We also use data on the services received by program group members, from either RAPTER or the program's management information system, to monitor the extent to which program group members are receiving services.

Statistical precision of estimates

The total sample size for each evaluation needs to be large enough to detect impacts of a size expected from the program. With these values in mind, the target sample size for all programs is 1,000 study participants—500 each in the program and comparison groups. The first program started enrollment in June 2021, and the last program started in May 2022. Enrollment will end for each program in June 2024.

Exhibit 11 shows the minimum detectable impacts by sample size. For the administrative data, we expect data for 95 percent of the sample, with five percent not reporting their Social Security number or reporting an invalid number. We expect about an 80 percent response rate for each of the participant follow-up surveys.

The minimum detectable impacts on earnings for a randomly-assigned sample of 1,000 are about $183 for average monthly earnings reported on the survey and $504 for quarterly earnings reported from administrative data. These minimum detectable impacts correspond to minimum detectable effect sizes of 0.18 for the survey sample and 0.16 for the administrative data sample. We also include power calculations for a sample size of 300, which could represent the size of a subgroup, or could reflect actual sample size obtained if one or more of the programs cannot recruit the full sample of study participants. The minimum detectable impacts would be about $334 (effect size = 0.32) for average survey monthly earnings and $920 (effect size = 0.30) for administrative data quarterly earnings, respectively.

Evidence reviews, such as the What Works Clearinghouse, consider effect sizes of 0.25 standard deviations or larger as substantively important (U.S. Department of Education 2020).

Exhibit 11. Minimum detectable impacts, by sample size, on earnings

Study sample (program and comparison)	Monthly earnings measured with survey data		Quarterly earnings measured with administrative data	
	Minimum detectable impacts in dollars	Minimum detectable impacts in effect sizes	Minimum detectable impacts in dollars	Minimum detectable impacts in effect sizes
1,000	$183	0.18	$504	0.16
300	$334	0.32	$920	0.30

Assumptions: individuals are randomly assigned; equal random assignment probabilities for program and comparison groups; $1,034 standard deviation of monthly earnings and $3,102 standard deviation of quarterly earnings; covariates explain 20 percent of the variation in the outcomes; response rate of 80 percent on the survey; match rate of 95 percent for the administrative data; two-tailed test, *p*-value of 0.05.

Impact study data sources and uses

We collect data from the following data sources to support the impact study: (1) a baseline survey; (2) two follow-up surveys; and (3) administrative data sources. This section discusses how each of these impact study data sources are collected and will be used for the impact studies.

Baseline survey

As noted previously, a short survey collects baseline data about all study participants before random assignment.

Use of baseline data

The baseline data will be used to conduct the following analyses for the impact study (in addition to its uses for the descriptive study):

- **Describe the characteristics of study participants and check that random assignment has created program and comparison groups with similar characteristics.** Descriptive analysis of baseline data will contextualize the research findings by identifying the demographic and socioeconomic characteristics of the research sample. We will also use these data to compare the baseline characteristics of the program and comparison groups to confirm that random assignment resulted in research groups that were initially similar along various dimensions.
- **Provide covariates for regression models.** Including covariates that are correlated with the outcome measure will improve the statistical precision of the impact estimates. These covariates will include baseline measures of key outcomes used in the impact analysis. These covariates will also include baseline characteristics that differ significantly between the program and comparison groups; these differences may emerge by chance despite the random assignment design.
- **Construct weights to adjust for survey nonresponse.** The nonresponse weights will adjust the data to be representative of all sample members, not just those who completed the survey or could be matched to an administrative record. We will calculate the weights by estimating, for each program separately, the probability of nonresponse for study participants as a function of their baseline characteristics. The baseline variables selected for use in the weighting analysis must be strongly correlated with whether participants responded to follow-up surveys. These variables could include factors associated with survey staff's ability to contact a study participant at follow-up, such as whether the participant had an email address at baseline. They could also include demographic and socioeconomic characteristics.
- **Support subgroup analysis.** We plan to conduct subgroup analyses to examine whether a program is more effective for some groups of participants than others. Subgroups may be defined based on participants' characteristics collected at baseline, including previous employment, educational attainment, and baseline measures of health and well-being.
- **Locate study participants for the follow-up surveys.** Detailed contact information is collected at baseline, including telephone numbers, addresses, and email addresses, to help locate participants to complete the follow-up surveys. We also collect detailed contact information for up to three relatives, friends, neighbors, and/or past employers, selected by participants, who may be able to help locate the participant if he/she moves.
- **Support "dosage" analysis and analysis of the mediating factors driving intervention impacts.** The analysis of the impacts of different "doses" or intensity of the receipt of services will benefit from data on factors related to a participant's propensity to receive program services, including participant demographic and socioeconomic characteristics.

The mediation analysis will require baseline data on factors such as a participant's employment history or baseline health status.

We collect most of the baseline data needed to conduct these analyses via the baseline survey. We will collect additional baseline data from administrative records, as discussed in greater detail in a later section.

Pre-testing survey

We conducted two rounds of pre-tests of the baseline survey with individuals similar to those served by the programs included in the NextGen Project. We conducted the first round of pre-tests before the programs had been selected; this pre-test included seven people who were receiving services from an organization in New Jersey that serves individuals experiencing poverty or homelessness. We then revised the survey and pre-tested it with another seven people who were served by two NextGen programs—Bridges and IPS-AJI. We selected these two programs specifically for the second pre-test effort because they serve populations with special considerations; the revised survey included unique questions tailored to them. Bridges serves young adults; IPS-AJI serves individuals with previous criminal justice involvement. For both rounds of pre-tests, we timed the interviews and used cognitive interviewing along with respondent and interviewer debriefings to assess respondents' understanding of the survey questions, identify improvements to the flow and structure of the instruments, and ensure that the survey length was as indicated. We updated the surveys based on the findings from both sets of pre-tests.

Data collected at baseline

The data collected at baseline through the baseline survey and RAPTER are shown in Exhibit 12. Not all of these items are collected for all programs—the survey instruments were tailored to include questions to be asked for specific evaluations.

Exhibit 12. Data collected at baseline

Domains and measures
Locating information. Name, date of birth, Social Security number, home telephone number, cell phone number, social media contact information, additional phone numbers, address, email address, contact information for additional contacts
Demographic and socioeconomic characteristics. Whether Hispanic, race, whether currently in high school, the highest degree or year of school have attended, primary language spoken at home, marital status, whether lives with a spouse or partner, number of adults with whom respondent lives, number of children with whom respondent lives, whether caring for someone in the household with a disability, housing status currently and during the past month, whether have been homeless in the last three months, whether currently receiving child support, whether expected to pay child support, whether received income or assistance in the past year from the following programs: SSI/SSDI, TANF, UI, worker's compensation, short-term disability, SNAP, Supplemental Nutrition Program for Women, Infants, and Children (WIC), housing voucher, veteran benefits, or Medicaid or Children's Health Insurance Program (CHIP), whether received SSA benefits currently because of a disability, whether received SSA benefits in the past year because of a disability, whether applied for Social Security disability benefits in the past five years, whether waiting for a decision of a disability application currently

Domains and measures
Social support. Number of people can count on to help, number of people can borrow $100 from, number of people can talk to for advice, whether respondent thinks most people can be trusted, whether respondent thinks most people in the neighborhood can be trusted
Employment status and history. Whether ever worked for pay, whether currently working for pay, months worked for pay during the past year, month and year when last worked for pay, amount paid before taxes and deductions at current or most recent job, number of hours worked per week at current or most recent job
Employment challenges. Whether a physical, mental or emotional condition limits the kind or amount of work one can do, whether the work limiting condition is related to COVID-19
Goal-related skills. Level of agreement on setting long-term employment goals and setting specific short-term goals, level of agreement on actions to achieve the goals
Criminal justice system involvement: Whether ever arrested, number of times convicted of a crime, number of times convicted of a felony, whether currently under supervision, type of crime charged for, whether incarcerated, and if so, total amount of time spent in incarceration and amount of time spent in last incarceration
Physical and mental health. Mental and physical health status as measured by the SF-12 Instrument (general health status, whether and to which extent health status limits types of activities), whether has a disability based on American Community Survey disability questions (deaf/serious difficulty hearing, blind/serious difficulty seeing, serious difficulty concentrating, remembering or making decisions, serious difficulty walking or climbing stairs, has difficulty dressing or bathing, having difficulty doing errands), ever treated for mental health condition, type of physical, mental, or emotional conditions that limited the respondent, mental health and distress status based questions adapted from K-6 Distress Scale or CESD-R, substance abuse disorders based on AUDIT-C questionnaire and DAST-10 scales

Note: The 12-Item Short Form Health Survey (SF-12) is a health-related quality-of-life questionnaire consisting of twelve questions that measure eight health domains to assess physical and mental health.

American Community Survey disability questions ask six disability types: hearing difficulty, vision difficulty, cognitive difficulty, ambulatory difficulty, self-care difficulty, and independent living difficulty. Respondents who report anyone of the six disability types are considered to have a disability.

The Kessler Psychological Distress Scale (K-6 scale) is a six-item self-report measure of psychological distress intended to be used as a quick tool to assess risk for serious mental illness in the general population.

Center for Epidemiologic Studies Depression Scale Revised (CESD-R) is a 20-item screening test for depression and depressive disorder.

The Alcohol Use Disorders Identification Test-Concise (AUDIT-C) questionnaire is a brief alcohol screening instrument that reliably identifies persons who are hazardous drinkers or have active alcohol use disorders (including alcohol abuse or dependence).

The Drug Abuse Screening Test (DAST-10) Scale is a 10-item brief screening tool that assesses drug use, not including alcohol or tobacco use in the past 12 months.

Administration of the baseline survey

The programs' intake staff were trained to administer the baseline survey to consenting study participants. Baseline data are stored on a secure data server. We monitor quality and consistency in the data collection by regularly reviewing the data, looking for patterns of missing data, and noting any other data quality issues.

Follow-up surveys

We collect data from two follow-up surveys of study participants, including both program and comparison group members.

Use of follow-up survey data

Data from the two follow-up surveys will be used for the following purposes:

- **Assess program effectiveness.** The responses to questions on the follow-up surveys will be used to develop outcome measures, which will be the basis of our assessment of program effectiveness. Not all questions on the follow-up surveys will be collected for all programs.
- **Describe the services received in the community and estimate the impact of the program on service receipt**. To interpret impact estimates, we need to understand the differences in the services received by the program and comparison groups. RAPTER and programs' management information systems provide information about services received by the program group from the program being evaluated, but generally do not provide information about services received from other providers in the community, or those received by the comparison group. The follow-up surveys collect information about services received by both study groups from the program being evaluated and other organizations in the community.
- **Examine the mechanisms through which a program operates.** The selected programs target key intermediate outcomes on the path to improved labor market, economic independence, and other outcomes. Important intermediate outcomes include receipt of health services, health status, and social support.
- **Assess perceptions of program usefulness**. We ask respondents in the program group to report what aspects of the program they found most useful and whether they attribute their subsequent outcomes to the program, potentially providing information about the components of the programs that led to impacts.
- **Locate study participants for the second follow-up survey.** The first follow-up survey collects detailed contact information, analogous to the information collected at baseline, to facilitate locating study participants for the second follow-up survey.

The first and second follow-up surveys are similar. Both collect information about service receipt and outcomes. The first follow-up survey asks respondents about the period between study enrollment and when the first follow-up survey is administered; the second follow-up survey asks respondents about the period between the administration of the two surveys. For those who respond to the second follow-up survey but not the first, the second follow-up survey asks respondents about the period between study enrollment and the time the second follow-up survey is being administered.

Length of follow-up period

In Bridges and Philly WINs, the surveys are administered approximately 9 and 21 months after study enrollment; the surveys for the IPS-AJI and Western Mass MOMS evaluations are administered approximately 6 and 18 months after study enrollment. The survey timing was based on when impacts were expected to emerge, based on each program's logic model.

Because evidence suggests that recall of jobs more than a year in the past is poor (Mastri et al. 2018), we did not want more than a year between study enrollment and the first follow-up survey or between the first and second follow-up surveys. The rationale for the length of the follow-up period for each program is described in the appendices.

Development and pre-testing of follow-up survey

Many questions on the follow-up surveys were sourced from existing validated instruments. Other questions came from scales frequently used in large-scale national surveys, such as the SF-12 Health Instrument, to assess health status. Content experts at Mathematica developed certain items. For example, we developed survey questions to measure confidence in the ability to seek employment for the Bridges evaluation based on the program's logic model.

The follow-up surveys include different questions for different evaluations. This is because situations and outcomes differ by evaluation. For example, for Bridges, which serves young adults with disabilities who are transitioning out of high school, to make the survey simple, we start by asking whether people are still in high school. For Western Mass MOMS, which focuses on reducing depressive symptoms, we ask a longer list of questions about depression symptoms.

Similar to pre-tests we conducted for the baseline survey, we conducted two rounds of pre-tests of the first follow-up survey with 11 individuals similar to those served by the programs included in the NextGen Project. The first round of pre-tests was conducted before the programs had been selected with four people in New Jersey who are like those in the focal populations. We then revised the survey and pre-tested it with seven people who were served by Bridges and IPS-AJI. For both rounds of pre-tests, we timed the interviews and used cognitive interviewing along with respondent and interviewer debriefings to assess respondents' understanding of the survey questions, identify improvements to the flow and structure of the instruments, and ensure that the survey length was as indicated. We updated the surveys based on the findings from both sets of pre-tests. Due to the similarity between the two follow-up survey questionnaires, we did not conduct a pre-test specifically for the second follow-up survey.

Data collected

The follow-up outcome measures collected on the surveys are shown in Exhibit 13. As discussed previously, not all of these outcome measures are collected for all programs—the surveys are tailored for each evaluation.

Exhibit 13. Data collected through the follow-up surveys

Domains and measures
Employment
Current employment status and formal employment history: Employment status, information for each job held from study enrollment through survey administration: begin and end date, wage rate, hours worked, earnings
Job quality: Type of job (regular, temporary, seasonal, contractor), schedule, fringe benefits, whether been promoted, expectations for advancement in current job, satisfaction with job
Access to work accommodations: Whether employer provided accommodations (special equipment, work schedule, task assignment, work environment, assistance from co-workers)

Domains and measures
Economic independence
Receipt of public assistance and other sources of income: Receipt of SSDI, SSI, TANF, unemployment insurance, SNAP, WIC, Medicaid or CHIP
Financial independence from family: Whether receive financial assistance from family and friends, and amount received
Perceived financial independence: Whether participants are confident that in five years' time they will earn enough to support themselves without financial help, that in five years' time they will be working at a paid job
Health
Health status: Whether participants have a physical, mental, or emotional condition limiting the kind or amount of work they can do, mental and physical health status as measured by the SF-12 scale, mental health and distress status measured by questions from K-6 Distress Scale or CESD-R, general happiness, substance abuse disorders (AUDIT-C, DAST-10 scales), opioid use for nonmedical reasons
Health insurance: Health insurance coverage
Economic well-being
Economic hardships: Experiences associated with economic hardship (could not pay the full amount of the rent or mortgage, was evicted for not paying the rent or mortgage, filed for bankruptcy, did not pay the full amount of the utility bill, had utilities turned off for not paying the bill, did not pay the full amount of child support payments because could not afford it, had telephone service disconnected because payments were not made, did not fill or postponed filling a prescription because could not afford it, did not pay other bills), whether have enough of the kinds of food
Housing status: Housing status (own, rent, live rent-free, live in shelter, unsheltered)
Service receipt
Receipt of employment services: Receipt of services since study enrollment, including the following: assessments, assistance with resume or application completion, assistance with job interview preparation, referrals for jobs, obtaining a drivers' license, obtaining other needed documentation, advice on presenting legal history and other legal help, assistance obtaining child care, assistance obtaining transportation, assistance obtaining clothing, tools, or other supplies for work, assistance meeting employers' COVID-19 requirements, assistance obtaining housing, information about receipt of benefits, advice on how to act at work and peer support, where received the most assistance from
Receipt of services related to physical or mental health: Receipt of assistance with substance use disorder since study enrollment and place received assistance with substance use disorder; receipt of mental health services, including place (hospital, doctor's office, etc.) received, number of times received; receipt of physical health services and how many times received; receipt of mental health services by respondent's children and place (hospital, doctor's office, etc.) children received services
Social support
Social support and trust: Number of people can turn to for advice, number of people can borrow $100 from, whether respondent thinks most people can be trusted, whether respondent thinks most people in the neighborhood can be trusted
Skill acquisition
Participation in and completion of education and training program: Participation in an education program, whether received additional degrees or certifications, participation in a training program, whether received training credentials
Goal-related skills: Level of agreement on setting long-term employment goals and setting specific short-term goals, level of agreement on actions to achieve the goals

Domains and measures
Employability skills: Eight-item scale about whether participants are confident in their abilities to conduct activities related to finding, applying for, progressing in, or keeping a job
Criminal justice system involvement
Criminal justice system involvement: Whether under supervision, number of times arrested since study enrollment, whether convicted since study enrollment, length of any incarceration, whether incarceration was due to supervision violations

Note: The 12-Item Short Form Health Survey (SF-12) is a health-related quality-of-life questionnaire consisting of twelve questions that measure eight health domains to assess physical and mental health.

The Kessler Psychological Distress Scale (K-6 scale) is a six-item self-report measure of psychological distress intended to be used as a quick tool to assess risk for serious mental illness in the general population.

Center for Epidemiologic Studies Depression Scale Revised (CESD-R) is a 20-item screening test for depression and depressive disorder.

The Alcohol Use Disorders Identification Test-Concise (AUDIT-C) questionnaire is a brief alcohol screening instrument that reliably identifies persons who are hazardous drinkers or have active alcohol use disorders (including alcohol abuse or dependence).

The Drug Abuse Screening Test (DAST-10) Scale is a 10-item brief screening tool that assesses drug use, not including alcohol or tobacco use in the past 12 months.

Data collection methods and expected response rates

Study participants can complete the follow-up surveys by either self-administering the survey via the web or completing the survey using computer-assisted telephone interviewing (CATI). We send email and physical mail notifications to study participants offering them the option of responding by web or phone. We send trained locators to attempt to meet nonresponders in person and offer them the use of a telephone to complete the interview.

We will attempt to complete both first and second follow-up surveys with the entire sample. Both surveys will take about 50 minutes on average to complete. We anticipate an 80 percent response rate on the follow-up surveys.

All interviews will take place in either English or Spanish, depending on the preference of the study participant. All interviewers were trained in interviewing techniques, the purpose of the study, the intent of each question, and in cross-cultural humility and understanding.

Administrative data from SSA, the National Directory of New Hires (NDNH), and other sources

We will collect administrative data to provide information about earnings and benefit receipt for all study participants. We will collect administrative data on criminal justice system involvement for study participants in the IPS-AJI evaluation. We plan to collect administrative data for up to two years after study enrollment. Because the consent form allows for collection of some administrative data over a follow-up period of up to 10 years, further collection of administrative data may occur.

Use of administrative data

We will use administrative records to describe the baseline characteristics of study participants, as well as assess the effectiveness of programs under evaluation in affecting employment, benefit receipt, and criminal justice system involvement outcomes. The data we collect vary by the evaluation, but may include the following:

- Employment, earnings, and new hires information
- Receipt of SSI and SSDI
- Receipt of TANF and SNAP
- Receipt of unemployment insurance
- Justice system involvement

Administrative data on employment, earnings, and benefit receipt offer several advantages over survey data. First, administrative data will be accessible for sample members for whom we have valid Social Security numbers even if they do not complete the follow-up survey. As a result, impact estimates for outcomes constructed using administrative data are less likely to be affected by nonresponse bias. Second, administrative earnings data are not subject to recall errors while evaluation surveys are likely vulnerable to this type of error (Bound et al. 2000; Mastri et al. 2018). For example, surveys ask study participants to recall their employment history early in the follow-up period. It is possible that respondents inadvertently fail to report jobs (forgetting or misremembering), especially those held early in the study period. Survey reports of benefit receipts are more likely to be subject to recall error and benefit underreporting (Meyer et al. 2009). The administrative records on SSI, SSDI, TANF, SNAP, and unemployment insurance provide more accurate information on benefit receipts. Further, compared to collecting survey data, it is relatively inexpensive to collect the administrative data.

But these types of administrative data have their own limitations. Employers may also underreport earnings to evade paying taxes, and often do so (Abraham et al. 2013; Moore et al. 2018). Because administrative earnings data sets are developed for purposes other than research, many include information on earnings from only a subset of jobs; this will likely lead to underreporting of total earnings, especially from jobs that people in our study sample are likely to hold in the gig economy (see, for instance, Abraham et al. 2019 and Katz and Krueger 2019). Administrative data do not include details of respondents' jobs. Furthermore, some types of administrative data—such as benefit receipt—are often limited to a local geographic area.

Administrative data on justice system involvement also has benefits and challenges. In addition to recall error, on surveys, respondents are prone to not report their arrests, convictions, or incarcerations even when they remember these events, leading to underreporting (Tourangeau and Yan 2007). However, administrative justice system data is fragmented, maintained by disparate public agencies. Its quality and completeness are unknown. Thus, it is unclear whether the administrative data provide more accurate information on criminal justice system involvement.

Administrative data sources and measures collected

NDNH. This database of earnings and employment information is maintained by ACF's Office of Child Support Services. NDNH data include quarterly earnings from all jobs covered by unemployment insurance. They do not include earnings from employment not covered by unemployment insurance, such as certain types of farm labor and work by independent contractor, or any jobs in which the employer does not report the earnings.

We will use NDNH data to provide information on earning history at the time of study enrollment and outcomes, such as consequent employment, earnings, becoming a new hire, and unemployment insurance receipt. Throughout the study enrollment period, we are conducting quarterly "input" data requests to preserve records that otherwise would be destroyed after the two-year NDNH data maintenance period. By using this approach, we will be able to collect NDNH data that cover the year and a half before study enrollment and two years after enrollment for all study participants, as well as for longer follow-up periods for those who entered the sample earlier in the enrollment period.

Administrative data from SSA. We will collect data on earnings and disability benefit receipt using several SSA data sources. The Master Earnings File (MEF) data capture earnings and self-reported income reported to the Internal Revenue Service and subject to Social Security taxes, including wages as well as self-reported income not included in NDNH. However, like the NDNH, it does not include data on under-the-table or informal employment that can be common among low-wage workers. We will use MEF data to provide information on earnings history and annual employment and earnings outcomes.[5] The MEF earnings data will cover five calendar years for most study participants—two years before study enrollment and three afterward.

Additional data will come from various SSA administrative files, which could include the 831 file, Structured Data Repository, Master Beneficiary Record, Payment History Update System, and Supplemental Security Record file. These data include SSI and SSDI application date, whether it is a SSI or SSDI application, decision date, decision level, whether granted a benefit, type of benefit received, primary disabling condition for the basis of the disability determination, benefits due/paid, and death information.

TANF and SNAP administrative data. TANF and SNAP data are available via state or county data warehouses. We will collect data for selected programs that serve a substantial proportion of people receiving or eligible for receiving such benefits and whose logic models include reduced need for these benefits as an outcome. We will use TANF and SNAP administrative records data to provide information on the amount of program benefits received. We will collect records that cover the year prior to study enrollment and two years after enrollment.

Administrative justice system data. Data on arrests, convictions, and incarcerations can be collected from some state and local agencies. These data can provide information about histories of arrests, convictions, sentences, and periods of incarceration for the IPS-AJI evaluation. We will take an exploratory approach to collecting these data. For instance, we

[5] The accessibility of MEF data is limited to researchers at SSA. We will work with SSA researchers to develop specifications for data analyses, and SSA will share aggregate output with the study team.

will explore the collection of the criminal justice data for one or two states first to analyze the differences between the administrative data and survey data at the first follow-up, and then determine the value of collecting them for all states for the second impact analysis.

Impact study analytic approach

This section outlines the analytic methods we will use to estimate the impacts of programs under the NextGen Project, describe exploratory analyses for additional research questions, and address potential methodological challenges.

Overview

The main impact estimates for all outcomes will be based on the evaluation's experimental design. With random assignment, the members of the groups should be on average similar in their characteristics before study enrollment. Our basic analytic approach is to compare the mean outcomes of members of the program and comparison groups after study enrollment. This approach will provide unbiased estimates of the impacts of the program.

To obtain more precise estimates, regression models will control for random differences in the baseline characteristics of program and comparison group members. In their simplest forms, these models can be expressed by the following equation:

(1) $$Y_i = \alpha + \beta X_i + \delta T_i + \varepsilon_i,$$

where Y_i is an outcome (such as earnings) for person *i*; α is a constant; X_i is a vector of baseline characteristics (such as gender, age, race/ethnicity); β is a vector of coefficient parameters for the extent to which baseline characteristics are predictive of the outcome; T_i is an indicator for whether person *i* received the program; δ represents the impact of the program; and ε_i is an error term. We will calculate standard errors using heteroskedasticity-robust standard errors that allow for the variance of the error term to differ by treatment status (Huber 1967; White 1980). For binary outcomes, we will use a linear probability model for the main specification.

We will estimate impacts separately for each program. For programs that have more than one geographic location participating in the evaluation, the main analysis will pool estimates across the locations and include an indicator for the locations in the regression model. This approach weights each participant of a program equally, rather than weighting location-level impacts equally. This is appropriate because each program is implementing a similar approach across different geographic locations and the approach improves the statistical precision of the impact estimates.

Reporting the findings

We will use two approaches: (1) a frequentist approach that reports statistical significance, and (2) a BASIE (BAyeSian Interpretation of Estimates) approach to calculate the probability that a program has particular effects.

Frequentist approach

For each impact estimate, we will report statistical significance based on *p*-values. The *p*-value reflects the probability of obtaining the observed impact estimate when the null hypothesis of no effect is true. We will deem impact estimates to be statistically significant if the associated *p*-value of the estimate falls below five percent based on a two-tailed hypothesis test (Exhibit 14). We will also note if the associated *p*-value falls between five and 10 percent, classifying these impacts as statistically significant at the 0.10 level.

Exhibit 14. Conventions for describing statistical significance

p-value of impact estimate	Symbol used to denote *p*-value	Description of impact estimate
$p < 0.01$	***	Statistically significant
$0.01 \leq p < 0.05$	**	Statistically significant
$0.05 \leq p < 0.10$	*	Statistically significant at the 0.10 level
$p \geq 0.10$	None	Not statistically significant

Bayesian approach

We plan to present findings using a Bayesian approach known as BASIE (Deke and Finucane 2019).[6] This will provide a probability that the true effect of the program is positive or greater than a specified amount. For example, we could draw conclusions about the likelihood that the impact was positive, such as "There is a 75 percent chance that the program had a positive effect on average monthly earnings." In addition, we could draw conclusions about the probability that the program had a large effect that readers are likely to regard as meaningful, such as: "There is a 50 percent chance that the program boosted average monthly earnings by $250 or more." The findings expressed in this way can be more helpful to practitioners and policymakers than just a conclusion that the program is probably effective or not. The Bayesian approach also guards against a misunderstanding about the meaning of statistical significance that can lead to serious misinterpretation of study findings—many people misinterpret statistical significance (*p*-value < 0.05) to mean that there is at most a 5 percent chance that the program had no effect rather than the correct conclusion that when the true effect is zero, there is a five percent chance that the impact estimate is statistically significant.

Additional analyses

We plan to conduct additional analyses: (1) estimating impacts by subgroup, (2) estimating impacts at different points in the distribution of outcomes, (3) mediation analysis, (4) estimating impacts on program group members who actually participate in the program (treatment on the treated [TOT]), (5) examining whether the impacts vary by the amount of program services received, and (6) conducting robustness checks.

[6] The components of BASIE draw on guidance from many sources (Gigerenzer and Hoffrage 1995; Gelman and Weakliem 2009; Gelman 2001, 2012, 2015a, 2015b, 2016; Gelman and Shalizi 2013).

Subgroup analysis

Identifying differences in impacts by groups of participants, determined by their characteristics at study enrollment, could help programs think through whether there is a need to revise or further target services.

We will estimate subgroup effects with the following specification:

(2) $$Y_i = \alpha + \beta X_i + \delta_1 T_i + \delta_2 G_i + \delta_3 T_i G_i + \varepsilon_i ,$$

where G_i is an indicator for whether person *i* is part of a subgroup and δ_3 represents the additional effect of a program for those in the subgroup. We will determine subgroups that align with each evaluation's logic model and base them on the population served. All subgroups will be measured using either baseline survey or administrative data.

For adequate statistical power, we will not estimate separate impacts for subgroups with fewer than 300 study participants. As shown in Exhibit 11, a sample of 300 study participants would provide a minimum detectable effect size of 0.29 for outcomes based on administrative records and 0.32 for those based on survey reports.

Examining impacts at different points of the distribution of outcomes

The impact analysis will be based on differences in the mean outcomes of program group members and comparison group members. This comparison provides a measure of the average impact of a program, but it does not tell us how the program affected the distribution of outcomes of interest. For example, it may be that the program had a different impact on earnings for those whose earnings would have been relatively high in the absence of the program than for those whose earnings would have been low in the absence of the program.

To learn more about program effects on the distribution of outcomes, we will examine differences between program and comparison group outcomes at different points in their distributions. We will use a multivariate statistical model to estimate these effects—an approach known as quantile regression analysis (Cook and Manning 2013).

Mediation analysis

We will consider conducting a mediation analysis, which could shed light on the mechanisms through which a program operates. For example, we might explore whether impacts on employment occur because the program improves the mental health of participants in Western Mass MOMS. To do so, we will decompose the overall effect of the intervention into a component due to the effect of the program on mental health and the estimated effect of mental health on the final outcomes of interest, such as employment, using a two-step procedure (Heckman et al. 2015; Kautz and Zanoni 2015).

Estimates of the impacts on those who participate in the program

Our main impact analysis will compare outcomes for all those assigned to the program group to those assigned to the comparison group and provide estimates of the "intent to treat" (ITT) impact. However, policymakers and program administrators are also interested in

estimates of the impact of the intervention on those who actually participated in the intervention—the treatment-on-the-treated or TOT impact. To estimate the TOT impact, we will apply the Bloom adjustment (Bloom 1984), which involves dividing the ITT estimate by the percentage of the program group who received any program services. This approach is valid if the members of the program group who did not participate in the program are unaffected by the program and no members of the comparison group receive services from the program.

Estimates of dosage effects

We might also explore how impacts vary by the intensity or dosage of services received by participants, or the types of services they receive. Specifically, we will use quasi-experimental propensity score matching methods to compare the outcomes of program participants who received a certain intensity of services to the outcomes of a subset of comparison group participants with similar background characteristics. We will first estimate the extent to which baseline characteristics predict program group members' participation intensity or take-up (Moore et al. 2012; Schochet and Burghardt 2007). We will then apply these predictors to members of both the program and comparison groups, and compare the outcomes of those in the program and comparison groups who are predicted to engage in services.

Methodological challenges

The evaluation may encounter methodological challenges, including missing data and accounting for the number of statistical tests in the impact analysis.

Missing data. Missing data could introduce bias in the impact estimates and reduce statistical power to detect program impacts. For missing baseline data, we will use dummy variable adjustment. This approach involves setting any missing baseline values to a single constant value and including indicator variables for missing values as additional covariates in the regression model. This approach is appropriate when the covariates are not correlated with the research group, as is the case in evaluations with a random assignment design (Deke and Puma 2013; Puma et al. 2009).

For missing outcome data, we will estimate regressions using weights to account for sample members who did not complete the follow-up survey or could not be matched to the administrative data because of missing or inaccurate Social Security numbers. The nonresponse weights will adjust the data to be representative of all sample members, not just those who completed the survey or could be matched to an administrative record. In addition, we will use imputation to address item nonresponse that affects the creation of survey outcomes.

Multiple hypothesis testing. The programs under study aim to influence a wide range of outcomes related to a number of domains. We are mindful that the probability of spuriously identifying impacts as statistically significant increases with the number of outcomes examined (Schochet 2009). For example, if 100 independent statistical tests are performed, with five percent set as the threshold for statistical significance, on average, five results will be statistically significant by chance alone even if the program had no impact. Furthermore,

this scenario has a 99.4 percent chance of at least one statistically significant result that is due to chance.

A key challenge in the impact evaluation will be to balance the need to cover the full set of outcomes that could be affected by the programs under study with the need to reduce the likelihood of generating multiple spurious program impacts. Deciding which outcomes to include in our assessment of program effectiveness will require careful consideration of the programs' goals. Key outcomes should be not only substantively important but also focused on areas in which the program is likely to have an impact. These outcomes will be selected based on the program logic model (see appendices for logic models).

We will develop formal procedures for identifying the main outcomes that indicate effectiveness for each program and document them in an analysis plan before we begin conducting the impact analysis. We will document these analytic decisions by registering the study at https://osf.io.

Comparing impacts to costs

We plan to conduct a benefit-cost analysis that compares the benefits arising from each program with its costs. This will help policymakers, practitioners, and participants decide whether to move forward with replicating, expanding, or participating in the program. This is also a way to put the magnitude of the impacts in perspective—impacts measured in dollar terms that are small compared to costs have different implications from impacts that are large compared to costs.

We will measure the dollar value of all the benefits and costs for which it is feasible to do so within the NextGen Project. Three important elements are relatively straightforward to measure: (1) the benefit of improved productivity and associated increases in taxes can be easily derived from the impact on earnings; (2) the benefit of a reduction in the use of public assistance programs (a cost to participants) can be derived from the impact on the dollar value of assistance received; and (3) the costs associated with providing program services. Other benefits—such as the benefit of improved mental health or reduction in criminal activity—can be estimated from the impacts on these outcomes but require information derived from other sources to estimate a dollar value of the benefits. Some benefits—such as the benefit of improved well-being not captured in other outcomes—cannot be measured but can be acknowledged. We will list all potential benefits, whether measured or not. Benefits that occur in the future will be discounted.

We will present the net benefits from four perspectives: (1) government agencies; (2) nongovernmental agencies and people; (3) participants; and (4) society as a whole. For each perspective, we will calculate net benefits per participant by subtracting the estimated costs per participant from the estimated benefits per participant.

References

Abraham, K.G., J. Haltiwanger, K. Sandusky, and J.R. Spletzer. "Exploring Differences in Employment Between Household and Establishment Data." *Journal of Labor Economics*, vol. 31, no. 1, 2013, pp. 129–172.

Abraham, K.G., J. Haltiwanger, K. Sandusky, and J. R. Spletzer. "The Rise of the Gig Economy: Fact or Fiction?" *AEA Papers and Proceedings*, vol. 109, 2019, pp. 357–361.

Avellar, S., R. Covington, Q. Moore, A. Patnaik, and A. Wu. "Parents and Children Together: Effects of Four Responsible Fatherhood Programs for Low-Income Parents." Washington, DC: Mathematica Policy Research, 2018.

Banerjee, S., P. Chatteriji, and K. Lahiri. "Effects of Psychiatric Disorders on Labor Market Outcomes: A Latent Variable Approach Using Multiple Clinical Indicators." *Health Economics*, vol. 26, no. 2, 2017, pp. 184–205. doi: 10.1002/hec.3286

Benjamini, Y., and Y. Hochberg. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society*, Series B (Methodological), vol. 57, no. 1, 1995.

Blase, K., and D. Fixsen. "Core Intervention Components: Identifying and Operationalizing What Makes Programs Work." Washington, DC: Office of the Assistant Secretary for Planning and Evaluation, 2013.

Bloom, H.S. "Accounting for No-Shows in Experimental Evaluation Designs." *Evaluation Review*, vol. 8, no. 2, 1984, pp. 225–246.

Bloom, D., P.J. Loprest, and S.R. Zedlewski. "TANF Recipients with Barriers to Employment." Temporary Assistance for Needy Families Program—Research Synthesis Brief No. 01. Washington, DC: Urban Institute, 2011.

Bond, G.R., D.R. Becker, and R.E. Drake. "Measurement of Fidelity of Implementation of Evidence-Based Practices: Case Example of the IPS Fidelity Scale." *Clinical Psychology: Science and Practice*, vol. 18, no. 2, 2011, pp. 126–141.

Braun, V., and V. Clarke. Thematic analysis. In H. Cooper (Ed.), APA handbook of research methods in psychology. Washington, DC: American Psychological Association, vol. 2, 2012, pp. 57-71.

Bronfenbrenner, U. "Developmental Ecology Through Space and Time: A Future Perspective." In *Examining Lives in Context: Perspectives on the Ecology of Human Development*, edited by P. Moen & G. H. Elder, Jr. Washington, DC: American Psychological Association, 1995.

Butler, D., J. Alson, D. Bloom, V. Deitch, A. Hill, J. Hsueh, E. Jacobs, S. Kim, R. McRoberts, and C. Redcross. "Enhanced Services for the Hard-to-Employ Demonstration and Evaluation Project: Final Results of the Hard-to-Employ Demonstration and Evaluation Project and Selected Sites from the Employment Retention and Advancement Project." OPRE Report #2012-08. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2012.

Cody, S., and C. Tuttle. "The Impact of Income Underreporting in CPS and SIPP on Microsimulation Models and Participating Rates." Washington, DC: Mathematica Policy Research, July 24, 2002.

Conrad K.J., C.I. Hultman, A.R. Pope, J.S. Lyons, W.C. Baxter, A.N. Daghestani, J.P. Lisiecki, P.L. Elbaum, M. McCarthy, and L.M. Manheim. "Case Managed Residential Care for Homeless Addicted Veterans: Results of a True Experiment." *Medical Care*, vol. 36, no. 1, 1998, pp. 40–53.

Cook, B. and W. Manning. "Thinking Beyond the Mean: A Practical Guide for Using Quantile Regression Methods for Health Services Research." Shangai Archives of Psychiatry, February 2013.

Cooper, H.M., L.V. Hedges, and J. Valentine (eds.). *The Handbook of Research Synthesis and Meta-Analysis.* Second edition. New York: Russell Sage Foundation, 2009.

Cox, D.R. *Analysis of Binary Data*. London: Chapman and Hall/CRC, 1970.

Cummings, D., and D. Bloom. 2020. "Can Subsidized Employment Programs Help Disadvantaged Job Seekers? A Synthesis of Findings from Evaluations of 13 Programs." OPRE Report #2020-23. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Damschroder, L., D. Aron, R. Keith, S. Kirsh, J. Alexander, and J. Lowery. "Fostering Implementation of Health Services Research Findings into Practice: A Consolidated Framework for Advancing Implementation Science." *Implementation Science*, vol. 4, no. 50, August 2009. Available at https://doi.org/10.1186/1748-5908-4-50. Accessed March 4, 2022.

Danziger, S., M. Corcoran, S. Danziger, C. Heflin, A. Kalil, J. Levine, D. Rosen, K. Seefeldt, K. Siefert, and R. Tolman. "Barriers to the Employment of Welfare Recipients." Institute for Research on Poverty Discussion Papers 1193-99. Madison, WI: University of Wisconsin Institute for Research on Poverty, n.d.

Deke, J., and M. Finucane. "Moving Beyond Statistical Significance: The BASIE (BAyeSian Interpretation of Estimates) Framework for Interpreting Findings from Impact Evaluations." OPRE Report #2019-35. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation, 2019.

Deke, J., and M. Puma. "Coping with Missing Data in Randomized Controlled Trials." Evaluation Technical Assistance Brief, no. 3. Washington, DC: Administration for Children and Families, Office of Adolescent Health, 2013.

Denzin, N.K. *Sociological Methods: A Sourcebook.* New York, NY: McGraw-Hill, 1978.

Deterding, N. and M. Waters. "Flexible Coding of In-depth Interviews: A Twenty-first-century Approach." *Sociological Methods & Research*, Vol. 50. Sage Publications, 2021.

DeVellis, R.F. *Scale Development: Theory and Applications.* Fourth edition. Los Angeles: Sage Publications, 2017.

Farrell, M., and K. Martinson. "The San Diego County Bridge to Employment in the Healthcare Industry Program: Implementation and Early Impact Report. Pathways for Advancing Careers and Education (PACE)." OPRE Report #2017-41, Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2017.

Frederick D.E., and T.J. VanderWeele. "Supported Employment: Meta-Analysis and Review of Randomized Controlled Trials of Individual Placement and Support." *PLoS One*, vol. 14, no. 2, 2019, e0212208. doi: 10.1371/journal.pone.0212208. PMID: 30785954; PMCID: PMC6382127

Gardiner, K., H. Rolston, D. Fein, and S. Cho. "Pima Community College Pathways to Healthcare Program: Implementation and Early Impact Report. Pathways for Advancing Careers and Education (PACE)." OPRE Report #2017-10, Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2017.

Gelman, A. "Ethics and Statistics: Ethics and the Statistical Use of Prior Information." *CHANCE*, vol. 25, no. 4, 2012, pp. 52–54.

Gelman, A. "Induction and Deduction in Bayesian Data Analysis." *Rationality, Markets and Morals*, vol. 2, 2001, pp. 67–78.

Gelman, A. "'The General Problem I Have with Noninformatively-Derived Bayesian Probabilities Is that They Tend to Be too Strong.'" Statistical Modeling, Causal Inference, and Social Science blog, 2015b. Available at http://andrewgelman.com/2015/05/01/general-problem-noninformatively-derived-bayesian-probabilities-tend-strong/. Accessed March 11, 2022.

Gelman, A. “Prior *Information*, not Prior *Belief*.” Statistical Modeling, Causal Inference, and Social Science blog, 2015a. Available at http://andrewgelman.com/2015/07/15/prior-information-not-prior-belief/. Accessed March 11, 2022.

Gelman, A. “What Is the ‘True Prior Distribution’? A Hard-Nosed Answer.” Statistical Modeling, Causal Inference, and Social Science blog, 2016. Available at http://andrewgelman.com/2016/04/23/what-is-the-true-prior-distribution-a-hard-nosed-answer/. Accessed March 11, 2022.

Gelman, A., J.B. Carlin, G.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis*. Third edition. Boca Raton, FL: CRC Press, 2013.

Gelman, A., and C.R. Shalizi. “Philosophy and the Practice of Bayesian Statistics.” *British Journal of Mathematical and Statistical Psychology*, vol. 66, no. 1, 2013, pp. 8–38. doi:10.1111/j.2044-8317.2011.02037x

Gelman, A., and D. Weakliem. “Of Beauty, Sex and Power.” *American Scientist*, vol. 97, no. 4, 2009, pp. 310–316.

Gigerenzer, G., and U. Hoffrage. “How to Improve Bayesian Reasoning Without Instruction: Frequency Formats.” *Psychological Review*, vol. 102, no. 4, 1995, pp. 684–704.

Glosser, A., C. Morrison, and D. Judkins. “Workforce Development Council of Seattle-King County Health Careers for All Program: Implementation and Early Impact Report. Pathways for Advancing Careers and Education (PACE).” OPRE Report #2017-106, Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2017.

Heckman, J., J. Humphries, and G. Veramendi. *The Causal Effects of Education on Earnings and Health*. Unpublished manuscript, 2015.

Hedges, L. “Distribution Theory for Glass’s Estimator of Effect Size and Related Estimators.” *Journal of Educational Statistics,* vol. 6, no. 2, summer 1981, pp. 107–128.

Hofmann, S., A. Asnaani, I. Vonk, A. Sawyer, and A. Fang. “The Efficacy of Cognitive Behavioral Therapy: A Review of Meta-Analyses.” *Cognitive Therapy and Research*, vol. 36, no. 5, 2012, pp. 427–440.

Hong, P., E. Gumz, S. Choi, B. Crawley, J.A. Cho. “Centering on Structural and Individual Employment Barriers for Human–Social Development”, *Social Development Issues,* vol. 43m, no. 1, 2022.
doi: https://doi.org/10.3998/sdi.1814

Huber, P.J. “The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions.” *Proceedings of the Fifth Berkeley Symposium on Math and Statistical Probability*, vol. 1: Statistics, 1967, pp. 221–233.

Katz, L.F., and A.B. Krueger. “The Rise and Nature of Alternative Work Arrangements in the United States, 1995–2015.” *ILR Review*, vol 72, no. 2, 2019, pp. 382–416.

Kautz, T., and W. Zanoni. *Measuring and Fostering Non-Cognitive Skills in Adolescents: Evidence from Chicago Public Schools and the OneGoal Program.* Unpublished manuscript, 2015.

Little, RJ. A, and D.B. Rubin. *Statistical Analysis with Missing Data*. 2. New York: Wiley; 2002.

Maguire, S., J. Freely, C. Clymer, M. Conway, and D. Schwartz. “Tuning in to Local Labor Markets: Findings from the Sectoral Employment Impact Study.” Philadelphia, PA: Public/Private Ventures, 2010.

Mann, D.R., D.C. Stapleton, and A. Porter. “Vermont’s Progressive Employment Program: A Preliminary Impact Analysis.” Princeton, NJ: Mathematica Policy Research, 2018.

Martinson, K., E. Copson, K. Gardiner, and D. Kitrosser. "Instituto del Progreso Latino's Carreras en Salud (Careers in Health) Program: Implementation and Early Impact Report. Pathways for Advancing Careers and Education (PACE)." OPRE Report #2018-06, Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2018.

Mastri, A., D. Rotz, and E.S. Hanno. "Comparing Job Training Impact Estimates Using Survey and Administrative Data." Washington, DC: Mathematica Policy Research, September 2018.

Maxwell, N.L., and D. Rotz. "Potential Assistance for Disadvantaged Workers: Employment Social Enterprises." *Journal of Labor Research,* vol 38, no. 2, 2017, pp. 129–141.

McConnell, S., I. Perez-Johnson, and J. Berk. "Proposal 9: Providing Disadvantaged Workers with Skills to Succeed in the Labor Market." Washington, DC: The Hamilton Project, Brookings Institution, n.d. Available at https://www.hamiltonproject.org/assets/legacy/files/downloads_and_links/disadvantaged_workers_skills_McConnell_Perez_Johnson_Berk.pdf. Accessed March 4, 2022.

Meyer, B.D., W.K. Mok, and J.X. Sullivan. "The Under-Reporting of Transfers in Household Surveys: Its Nature and Consequences." Working Paper 15181. Cambridge, MA: National Bureau of Economic Research, 2009.

Moore, Q., I. Perez-Johnson, and R. Santillano. "Decomposing Differences in Impacts on Survey and Administrative-Measured Earnings from a Job-Training Voucher Experiment." *Evaluation Review*, vol. 42, nos. 4–6, 2018, pp. 515–549.

Moore, Q., R.G. Wood, A. Clarkwest, A. Killewald, and S. Monahan. "The Building Strong Families Project: The Long-Term Effects of Building Strong Families: A Relationship Skills Education Program for Unmarried Parents, Technical Supplement." OPRE Report #2012-28C. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2012.

Morgenstern, J., A. Hogue, and J. Mckay. "Does Coordinated Care Management Improve Employment for Substance-Using Welfare Recipients?" *Journal of Studies on Alcohol and Drugs*, vol. 70, no. 6, 2009, pp. 955–963.

Mullainathan, S., and E. Shafir. Scarcity: Why having too little means so much. Henry Holt, 2013.

Orr, L.L. *Social Experiments: Evaluating Public Programs with Experimental Methods*. Thousand Oaks, CA: Sage, 1999.

Peck, L.R., A. Werner, E. Harvill, D. Litwok, S. Moulton, A.R. Fountain, and G. Locke. "Health Profession Opportunity Grants (HPOG 1.0) Impact Study Interim Report: Program Implementation and Short-Term Impacts." OPRE Report #2018-16. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2018.

Puma, M.J., R.B. Olsen, S.H. Bell, and C. Price. "What to Do When Data Are Missing in Group Randomized Controlled Trials." NCEE 2009-0049. Washington, DC: National Center for Education Evaluation and Regional Assistance, 2009.

Raphael, S. "The New Scarlet Letter? Negotiating the U.S. Labor Market with a Criminal Record." Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 2014.

Reed, D., A. Yung-Hsu Liu, R. Kleinman, A. Mastri, D. Reed, S. Sattar, and J. Ziegler. "An Effectiveness Assessment and Cost-Benefit Analysis of Registered Apprenticeships in 10 States." Oakland, CA: Mathematica Policy Research, 2012.

Rolston, H., E. Copson, and K. Gardiner. "Valley Initiative for Development and Advancement: Implementation and Early Impact Report. Pathways for Advancing Careers and Education (PACE)." OPRE Report #2017-83, Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2017.

Rotz, D., N.L. Maxwell, and A. Dunn. "Economic Self-Sufficiency and Life Stability One Year After Starting a Social Enterprise Job." Report submitted to REDF. Oakland, CA: Mathematica Policy Research, January 2015.

Sattar, S., M.A. Anderson, K. Chesnut, and V. Sotelo Munoz. "An Unprecedented Crisis: The WeCARE Program's Experience Serving People with Mental and Physical Health Challenges during a Pandemic." OPRE Report #2023-036. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2022.

Schochet, P.Z. "An Approach for Addressing the Multiple Testing Problem in Social Policy Impact Evaluations." *Evaluation Review*, vol. 33, no. 6, 2009, pp. 539–567.

Schochet, P.Z., and J. Burghardt. "Using Propensity Scoring to Estimate Program-Related Subgroup Impacts in Experimental Program Evaluations." *Evaluation Review*, vol. 31, no. 2, April 2007, pp. 95–120.

Smith M.V., L.S. Callinan, and M. Ciarlegio. "Can Treating Maternal Depression with Cognitive Behavioral Therapy Increase Employment among Low-Income Mothers? A Pilot Study." *Journal of Health Care for the Poor and Underserved*, 2021. Under review.

Smith, M.V., L.S. Callinan, C.S. Posner, S.C. Holmes, and R. Ebling. "Improving Maternal Mental Health as a Pathway to Economic Mobility in the TANF System." *Psychiatric Services*, 2021, forthcoming.

Treskon, L.. "What Works for Disconnected Young People: A Scan of the Evidence." New York, NY: MDRC, 2016.

Tourangeau, R, Yan, T." Sensitive Questions in Surveys." *Psychological Bulletin*, vol. 133, no. 5, 2007, pp. 859–883. doi:10.1037/0033-2909.133.5.859

U.S. Department of Education. WWC *Procedures and Standards Handbook.* Washington, DC: Institute for Education Sciences, 2020. Available at https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf. Accessed March 11, 2022

Vollmer, L., A. Mastri, A. Maccarone, and E. Sama-Miller. "The Right Tool for the Job: A Meta-Regression of Employment Strategies' Effects on Different Outcomes." OPRE Report #2017-40-A. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2017.

White, H.. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica*, vol. 48, no. 4, 1980, pp. 817–838.

Wu, April Yanyuan, Owen Schochet, Kristen Joyce, and Nicardo McInnis."Design and Analysis Plan for the Impact Study of Work Success," OPRE Report #2023-212, Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2023.

Ziegler, J.. "Sector Strategies: Aligning the Skills of the Workforce with the Needs of Employers." Mathematica Issue Brief. Oakland, CA: Mathematica Policy Research, November 2015. Available at https://www.mathematica-mpr.com/our-publications-and-findings/publications/sector-strategies-aligning-the-skills-of-the-workforce-with-the-needs-of-employers. Accessed March 11, 2022.

Appendix A: Further information about the design of the evaluation of Bridges from School to Work

Bridges from School to Work (Bridges) provides employment services to young adults (ages 17–24) with disabilities who are transitioning out of high school. Bridges' mission is "transforming the lives of young adults through the power of a job." The program has been operating for over 30 years and is located in 12 urban areas across the United States. It aims to meet the needs of both local employers and young adults with disabilities. Each participating young adult works with a Bridges staff member—called an employment specialist (ES)—who provides intensive and customized one-on-one support to help the young adult become ready for employment. To identify suitable job matches for participants, ESs use a strengths-based approach focused on the young adults' skills, interests, and abilities rather than a deficit-oriented framing focused on their disabilities. Staff members also assist participants during their job search and onboarding, and provide ongoing retention support for up to a year after they are placed in a job.

The evaluation of Bridges will provide more evidence than is currently available on the impact of the Bridges model. The evaluation is occurring in half of Bridges' locations.[7] Bridges leadership selected these locations for the NextGen Project because they had opportunities to expand into new parts of metro areas with an established Bridges presence and because they had experienced staff who knew the Bridges model well.

This appendix provides information about why Bridges was selected for evaluation and the program and its logic model. Supplementing the discussion in Chapters 3-4 of the report, it then discusses the formative evaluation that occurred and how the descriptive, cost, and impact evaluations are tailored for the Bridges evaluation.

Rationale for evaluation

Young adults with disabilities have employment rates that are substantially lower than those of their peers without disabilities (Wagner et al. 2005; Newman et al. 2009; Liu et al. 2018). For example, youth and young adults with disabilities (ages 14–24) are about 15 percent less likely to be employed than their peers without disabilities (Cheng & Richardson 2023). Without supports to encourage employment and independence, young adults with disabilities may experience poverty and high rates of public assistance receipt (Davies et al. 2009; Rupp et al. 2015). However, research has shown that young adults with disabilities can find and keep jobs if they have the opportunity and necessary supports (Martinez 2013). Researchers have also found that the opportunity to work during the secondary school years increases the likelihood of employment after the young adults leave school (Test et al. 2009; Carter et al. 2012).

Prior evidence suggests that Bridges may improve employment outcomes and reduce the need for Supplemental Security Income (SSI) for young adults transitioning to adulthood. Bridges' data over the past two decades show that enrollees have employment rates that are higher than other youth with disabilities (Fabian et al. 1998; Luecking and Fabian 2000; Fabian 2007; Gold et al. 2013). A nonexperimental study revealed that Bridges participation is associated with higher earnings and

[7] Most of the referrals for NextGen will occur through school districts, who prohibit us from identifying their participation or that of their students in our reports. For that reason, we do not list the specific study sites, to avoid the possibility of identifying the districts in some Bridges locations.

reduced receipt of federal disability benefits for participants through at least age 30 (Hemmeter et al. 2015).

Overview of program and its logic model

This section provides a brief overview of Bridges, what the program expects to change in the lives of the young adults it serves, and how it expects staff to go about changing those aspects of participants' lives (Exhibit A.1).

Participant referral sources

In consultation with school staff, Bridges primarily recruits young adults from Title I high schools in urban areas, which predominantly serve students from families with low incomes. For the NextGen Project, Bridges expanded into schools and school districts that were not previously referring young adults to Bridges. The intent of this decision was to preserve the availability of services in schools where Bridges staff were already providing services, and no young adult was denied services to which they otherwise would have been eligible in absence of the study.

Referred young adults are typically students in their last year of high school, though Bridges will work with participants in other grades provided they meet the program's age eligibility requirements. Less commonly, Bridges receives referrals from out-of-school referral sources including recent graduates, word-of-mouth, other community service providers, local community colleges, or online interest forms.

Eligibility criteria

To be eligible to participate in the NextGen evaluation of Bridges, the young adults must:

1. Be ages 17–24.
2. Have an Individualized Education Program (IEP), Section 504 plan, or other documentation of a disability from a medical professional.
3. Be interested in and able to work in competitive employment.
4. Be able to get to and from work.

The young adult needs to consent to participate in the evaluation and, if younger than age 18, their parent or guardian must also consent for them to participate in the evaluation.

Participant needs and resources

Bridges participants are interested in the program because of the services it provides related to employment; most participants do not have access to other similar services because (1) programs are geared to young adults with more significant disabilities than the typical Bridges participant, (2) programs are geared to disadvantaged youth without disabilities, or (3) programs such as state vocational rehabilitation services are relatively light touch and do not sufficiently address participants' needs.

Exhibit A.1. Bridges from School to Work logic model

Participant needs and program resources	Program activities and services	Short-term outcomes	Long-term outcomes
Participant needs • Work readiness and employability skills training • Ongoing support during the transition from school **Program resources** • Partnerships with school districts, schools, and teachers • Caseload size that allows for intensive and individualized supports • Ability to meet participants in the community, including physical space, technology, and reliable transportation • Knowledge of local employers' hiring needs • Well-trained staff, including ESs • Materials to support participants, including a curated set of materials to cover pre-employment through post-placement needs • Case management system to track progress toward employment and retention milestones	• In-depth assessment of participants' skills, interests, and abilities • ES as the single point of contact with frequent and sustained participant interaction • Customized job readiness skills training, job search support, and job retention services • Dual focus of participant and employer needs; outreach to new employer partners, assessing hiring needs of existing partners, and routine communication with supervisors where participants work • Performance incentives that prioritize job retention along with job placement	• Participants are more prepared for work, including having documents needed for work, and demonstrate work-related skills such as interpersonal skills, resume creation, job search and application, and getting to and from work • Employers find Bridges candidates well-prepared, want to hire more participants • Participants start working and gain experience in work based on their skills, interests, and abilities (meant to lead to longer-term success), not their disabilities • Participants stay on the job over many months, meeting Bridges' performance milestones • Participants have increased confidence and self-advocacy skills in work-related activities	• Sustained employment for participants and former participants • Promotion and/or advancement, within or across employers • Increased self-sufficiency and reduced need for financial assistance from family and other supports, as measured by milestones such as moving out of the family home or purchasing a car • Reduced need for public benefits, including applications to and receipt of SSI and SSDI
Community context • Participants interested in work with schedules that support work • Availability of entry-level jobs • Availability of staff with many competencies needed to be ESs		• Family and community support for employment • Availability and accessibility of transportation that is aligned to local employment opportunities	

The young adults that Bridges serves all have a range of disabilities and have generally received special education services at school (via their Individualized Education Program [IEP] and/or 504 plan). Participants are young adults in their late teens and early 20s; most primarily communicate by text message. Additionally, some participants come from families in which adults are not working, and in some cases believe (due to their own perceptions, or those of their families) that they may not succeed at work due to their disabilities. They also may have other social services needs related to housing instability, mental health needs, food insecurity, caregiving requirements, and other needs related to living with low incomes.

Bridges partners with schools to recruit high schoolers eligible for the program and works closely with teachers and other school staff to deliver employment-related training and other services during the school day. Bridges remains in close contact with teachers to ensure that program participants are on track at school and troubleshoots any challenges with teachers.

Many Bridges participants need assistance preparing for work. These needs include developing a resume, learning workplace expectations, obtaining interview clothes, and obtaining the documents needed to work such as a Social Security card or state identification card. They also need assistance applying for work, including finding job opportunities, filling out applications, completing employer assessments, taking public transportation to and from work, and polishing interview skills.

Once participants start working, they often need help with job onboarding activities and paperwork, and with general skills required for work such as requesting schedule changes, depositing paychecks, requesting workplace accommodations, interacting with coworkers, and responding to managers. Participants may also need support to change jobs following termination or resignation.

The NextGen Project pays for the Bridges services provided to NextGen study participants in the program group.

Activities conducted and services provided

Bridges focuses its service delivery on young adults, but considers its model to be employer-driven, meaning that it seeks to prepare participants to meet the hiring needs of local employers. ESs develop relationships with local employers to assess their needs, prepare participants to succeed in available jobs, and find qualified Bridges participants for open positions when possible.

Bridges provides individualized support to each participant based on their needs. ESs begin to work with participants before formally enrolling them into the program's caseload, then support participants on the caseload for up to a year after job placement. This includes supporting the participant in finding a new job, if needed—whether due to the participant's interests, poor performance on the job, or desire to advance. In those situations, ongoing support from ESs may last over a year after the first job placement.

The core components as determined by Bridges fall into five categories:

- **Focus on participants' skills, interests, and abilities—not their disabilities.** Bridges aims to equip participants to be well-prepared to find a job and succeed at work and seeks to match participants to positions that are well-suited to their skills and strengths. Bridges does not present participants to employers as students with disabilities, nor are disabilities typically discussed as part of the hiring process.

- **A single point of contact for each participant.** Each participant works with the same ES from their first interaction with the program, with no handoffs to other staff for the duration of the program. ESs maintain routine and frequent communication with the participant, at least weekly initially, then tapering to monthly once the participant is steadily employed. The goal is for ESs to become a trusted resource for participants. ESs also work with participants' teachers, parent(s)/guardian(s), and employer as needed.
- **Intensive supports and services spanning from pre-employment to job search to onboarding and retention.** ESs tailor services to participants across the continuum from pre-employment through retention. They routinely assist participants with preparing a resume, practicing communication skills, and preparing for and securing interviews. ESs assist participants with their job search—they might help them complete online applications, train them to use public transportation to get to and from work, and drive them to job interviews.

  After participants are hired, the ES assists them with new hire paperwork, background checks, and securing work uniforms. The ES checks in weekly to offer support and encouragement, offers on-site job coaching as needed, and advises on skill building for retention and advancement. In addition, the ES checks in with managers and co-workers to assess progress and correct any performance issues. ESs work with participants for up to a year after placement as needed and will help them identify another job if the current one is not the right fit.
- **Provision of services where participants live and work.** The Bridges model is centered on ESs making themselves available in the community. ESs are expected to meet participants where they live, work, and go to school, rather than making participants come to a central program location. ESs are also permitted to accompany participants on public transit or transport participants to employment-related activities in the ES's personal vehicle. In addition, ESs can be reimbursed for purchasing employment-related items for the participant, ranging from meals between interviews, to interview clothes, to assistance with obtaining required job certifications.
- **Maintaining dual focus on participants and employers.** ESs focus on matching the needs of local employers with job candidates Bridges prepares for work. ESs maintain strong relationships with local employers and routinely assess employers' hiring needs.
- **Well-defined performance metrics for staff that incentivize helping participants achieve and sustain employment.** ESs aim to enroll at least 20 participants into the program each year. Of those 20, 16 must be placed in a job, 12 must achieve 90 days employed, and 10 must achieve 180 days on the job (either at the original or subsequent placement). Achieving these milestones is part of an ES's performance review and factors into their annual compensation.

Expected outcomes

Bridges core services are expected to lead to five overarching short-term outcomes for participants:

- **Participants who are prepared to work.** ESs make sure that young adults have the documents necessary for employment; a polished resume; interpersonal skills, etiquette, and grooming habits suitable for the workplace; clothes suitable for interviewing; and travel training.
- **Participants who know how to search for jobs suitable to them.** ESs aim to equip young adults with knowledge about jobs suited to their skills and interests as well as their work experience. They want participants to consider job hours and transit when considering available options. ESs also train participants to look for jobs, fill out online applications, pass employer assessments, and follow up on submitted applications.

- **Participants who start working and stay employed.** Bridges aims to not only help participants start working, but also help them remain steadily employed. Bridges will continue to work with participants after resignation or termination to start a new job and sustain work, if the participants are interested in doing so.
- **Participants who improve their work-related self-advocacy skills, confidence, and independent decision-making.** Bridges aims for participants to know how to navigate situations at work, with increasing independence over time.
- **Participants who meet local employers' needs.** Bridges aims to prepare participants to be reliable and effective employees. When employers recognize the value of the Bridges program, they reach out to Bridges to meet hiring needs in the future.

Longer term, Bridges aims for participants to advance at work, either with their original employer or in a new job; increase their economic independence through increased earnings and reduced need for public benefits; and be role models in their communities and stewards of the Bridges program.

Formative evaluation

As we prepared Bridges to begin the random assignment evaluation, we identified a need to standardize how ESs documented the services they provided to their participants. In particular, we identified gaps and inconsistencies in how Bridges staff entered service data into ClientTrack, their management information system. Bridges leadership had already identified inconsistencies in how staff entered data across Bridges' 12 offices. For the evaluation, it was essential that the data entry on service receipt be consistent. The NextGen team analyzed an export of program year 2019 data and facilitated a series of meetings with Bridges staff to determine how to streamline the existing service categories (particularly those not used often) and improve consistency of data entry. Working with Bridges staff, the NextGen team proposed a smaller set of service categories and documentation for each one, pilot tested them with a small group of Bridges staff, refined them based on feedback, and tested them again. The program rolled out the revised service categories and documentation across the organization before the study was launched.

Descriptive and cost evaluations

Exhibit A.2 summarizes the evaluation-specific decisions made about the data collection for the descriptive and cost evaluations of Bridges:

Exhibit A.2. Data collection sources and methods for the Bridges descriptive and cost evaluations

Data collection	Respondents	Mode	Timing
Staff and leadership survey	13 ESs in offices participating in the NextGen Project along with 10 program leaders including directors in participating locations and in Bridges headquarters	Web	Fall 2022
Case reviews	Three cases served by different ESs across Bridges offices participating in the NextGen Project	Virtual	Winter 2022

Data collection	Respondents	Mode	Timing
In-depth staff interviews	12 ESs in offices participating in the NextGen Project along with 10 program leaders including directors in participating locations and in Bridges headquarters	Virtual	Spring 2023
Employer interviews	Three employers who hired at least one study participant	Telephone	Spring 2023
In-depth interviews with participants	17 participants who meaningfully engaged with Bridges after random assignment	Telephone, virtual	Spring 2023
Management information system (Program system/RAPTER)	Program group members who had been in the study for at least six months	Web-based	Data from August 2021 through descriptive study report final draft
Cost collection	Bridges central office staff	Telephone conversations and emails	Summer and fall 2023

For the cost study, we used a cost study workbook in Excel to collect data on the costs of implementing Bridges. We sent the tailored workbook to Bridges and held telephone conversations with Bridges central office staff in summer and fall 2023 to further explain the data request. We focused on costs incurred during calendar year 2022 because the period aligns with Bridges' fiscal year. By that time, most interruptions in service delivery due to the COVID-19 pandemic had subsided, other than a short disruption early in the year during the Omicron wave.

Impact evaluation

The impact evaluation is tailored to the evaluation of Bridges in three main ways: (1) random assignment procedures, (2) the services offered to the comparison group members, and (3) the data collected on the surveys.

Random assignment procedures

Random assignment occurs after the ES determines the young adult is eligible for Bridges based on the criteria listed above, obtains consent from the young adult and from parent/guardians if necessary, and the participant has completed the baseline survey.

Because Bridges eligibility criteria involve a subjective assessment of interest in work and ability to work independently, ESs can meet once or twice individually with participants to make that determination before enrolling them in the study. During those meetings, they are instructed not to provide any employment-related services.

Services offered to the comparison group

Bridges gives members of the comparison group a list of local resources to support them in their employment efforts as well as a $40 gift card.[8] Comparison group members have no further contact with Bridges after study enrollment. The school may connect members of both the program and comparison groups to auxiliary services, including referrals to employment programs in the

[8] The Bridges program paid for these gift cards.

community, social service agencies, physical or mental health services, or other public assistance such as SSI. Participants in both groups are also able to access any other employment supports available to them in the community, including through the state Vocational Rehabilitation program.

Data collection

The data collection for Bridges differs from the other programs in the length of the follow-up period and the tailoring of the baseline and follow-up surveys.

- **Follow-up period**. The follow-up surveys occur at nine months after random assignment and then again at 21 months after random assignment. Bridges works with participants for up to a year after they start work, with most of the concentrated effort in the months immediately after random assignment. The timing of the follow-up surveys allows for respondents to have had substantial experience with the program for nine months, and for the evaluation to observe the longer-term effects of the program one year later.
- **Tailoring the baseline and follow-up survey data collection**. The baseline and follow-up surveys were tailored for young adults, many of whom have limited work histories and may not be aware of their family's financial situations. In addition, given concerns about reporting accuracy, the surveys do not ask Bridges participants questions related to household income, receipt of public assistance or other benefits, or past justice involvement. The main differences are shown in Exhibit A.3.

Exhibit A.3. Tailoring of the baseline and follow-up surveys for Bridges

Change	Baseline	Follow-up	Rationale
Added a question to ask whether participants are currently in high school	✓		Bridges participants are typically, though not always, in high school when recruited
Omitted questions about household receipt of public assistance	✓	✓	Bridges participants are commonly teenagers, who may not be fully aware of the household's financial status
Modified response options for living arrangements to include an option about living with parents	✓	✓	Bridges participants are commonly teenagers who still reside in their childhood home
Trimmed the number of questions related to past work and justice involvement	✓	✓	Because of their age, Bridges participants typically have limited work history and limited involvement with the justice system

References

Carter, E.W., D. Austin, and A.A. Trainor. (2012). "Predictors of postschool employment outcomes for young adults with severe disabilities." *Journal of Disability Policy Studies,* vol. 23, no. 1, 2012, pp. 50–63.

Cheng, L., & Richardson, K. (2023). The 2023 youth transition report: Outcomes for youth and young adults with disabilities. Washington, D.C.: Institute for Educational Leadership.

Davies, P.S., K. Rupp, and D. Wittenburg. "A life-cycle perspective on the transition to adulthood among children receiving Supplemental Security Income payments." *Journal of Vocational Rehabilitation,* vol. 30, no. 3, 2009, pp. 133–151.

Fabian, E.S. "Urban youth with disabilities: Factors affecting transition employment." *Rehabilitation Counseling Bulletin,* vol. 50, no. 3, 2007, pp. 130–138.

Fabian, E.S., R.W. Lent and S.P. Willis. "Predicting Work Transition Outcomes for Students With Disabilities: Implications for Counselors." *Journal of Counseling & Development,* vol. 76, no. 3, 1998, pp. 311–316.

Gold, P.B., E.S. Fabian, and R.G. Luecking. "Job acquisition by urban youth with disabilities transitioning from school to work." *Rehabilitation Counseling Bulletin,* vol. 57, no. 1, 2013, pp. 31–45.

Hemmeter, J., M. Donovan, J. Cobb, and T. Asbury. "Long term earnings and disability program participation outcomes of the Bridges transition program." *Journal of Vocational Rehabilitation,* vol. 42, no. 1, 2015, pp. 1–15.

Liu, A.Y., J. Lacoe, S. Lipscomb, J. Haimson, D.R. Johnson, and M.L. Thurlow. "Preparing for life after high school: The characteristics and experiences of youth in special education. Findings from the National Longitudinal Transition Study 2012. Volume 3: Comparisons over time." (Full report) (NCEE 2018-4007) 2018, p. 210. Retrieved from U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance website: https://files.eric.ed.gov/fulltext/ED580934.pdf

Luecking, R.G., and E.S. Fabian. "Paid Internships and Employment Success for Youth in Transition." *Career Development for Exceptional Individuals,* vol. 23, no. 2, 2000, pp. 205–221.

Martinez, K. "Integrated employment, EmploymentFirst, and U.S. federal policy." *Journal of Vocational Rehabilitation,* vol. 38, no. 3, 2013, pp. 165–168.

Newman, L., M. Wagner, R. Cameto, and A.M. Knokey. "The Post-High School Outcomes of Youth With Disabilities up to 4 Years After High School: A Report From the National Longitudinal Transition Study-2 (NLTS2). NCSER 2009-3017." 2009. Retrieved from National Center for Special Education Research website: https://eric.ed.gov/?id=ED505448

Rupp, K., J. Hemmeter, and P.S. Davies. "Longitudinal Patterns of Disability Program Participation and Mortality Across Childhood SSI Award Cohorts (SSRN Scholarly Paper No. ID 2564689)." 2015. Retrieved from Social Science Research Network website: https://papers.ssrn.com/abstract=2564689

Test, D.W., V.L. Mazzotti, A. L. Mustian, C.H. Fowler, L. Kortering, and P. Kohler. "Evidence-Based Secondary Transition Predictors for Improving Postschool Outcomes for Students With Disabilities." *Career Development for Exceptional Individuals,* vol. 32, no. 3, 2009, pp. 160–181.

Wagner, M., L. Newman, R. Cameto, and P. Levine. "Changes over Time in the Early Postschool Outcomes of Youth with Disabilities. A Report of Findings from the National Longitudinal Transition Study (NLTS) and the National Longitudinal Transition Study-2 (NLTS2)." 2005. Retrieved from National Center for Special Education Research website: https://eric.ed.gov/?id=ED494920

Appendix B: Further information about the design of the evaluation of IPS for Adults with Justice Involvement

Individual Placement and Support for Adults with Justice Involvement (IPS-AJI) serves adults who have recently been involved in the criminal justice system and have a mental health diagnosis. The program is based on the IPS model, which was designed to help people with serious mental illness find and keep work at competitive jobs of their choosing. Six mental health centers in the Midwest or South are implementing IPS-AJI:

- Case Management Incorporated (CMI) in Memphis, Tennessee
- Central Oklahoma Community Mental Health Center (COCMHC)
- GRAND Mental Health (GRAND) in northeastern Oklahoma
- Pee Dee Mental Health (Pee Dee) in Florence, South Carolina
- Transitions Mental Health Services (Transitions) in Moline, Illinois, and Davenport, Iowa
- HOPE Community Services (HOPE) in Oklahoma City, Oklahoma

The core of the IPS model is the combination of mental health services with rapid job search and individualized pre-employment and follow-along support. An important component of IPS is the development of relationships between program staff and employers. IPS-AJI differs from the regular IPS model in that it only serves adults who have recently been involved in the criminal justice system and the adults do not need to have a diagnosis of serious mental health illness (such as schizophrenia spectrum disorder, bipolar, or depression). Following the IPS model, adults are not excluded based on the nature of their justice system involvement.

The IPS Employment Center, whose predecessor developed IPS, as well as state IPS trainers provide technical assistance, program manuals, and fidelity reviews to support implementation of IPS-AJI.

This appendix provides information about why IPS-AJI was selected for evaluation and the program and its logic model. Supplementing the discussion in Chapters 3–4 of the report, it then discusses how the descriptive, cost, and impact evaluations are tailored for the IPS-AJI evaluation.

Rationale for evaluation

The Dartmouth Psychiatric Research Center (which became the IPS Employment Center), in collaboration with clients and other experts, developed IPS in the late 1980s in response to discontent among people with severe mental illness with the existing employment program models. Since then, at least 1,000 IPS programs have been implemented in the United States in addition to programs in 21 other nations (Drake et al. 2020; Bond 2022).

Extensive rigorous evidence from randomized controlled trials demonstrates the efficacy of IPS for improving employment outcomes of individuals with severe mental illness (Frederick and VanderWeele 2019). Frederick and VanderWeele (2019) contains a meta-analysis of 25 randomized controlled trials, two follow-up studies that extended the period of observation of a previous trial, and three secondary analyses on previous trials.

While people with justice system involvement may not have diagnoses of serious mental illness, mental health issues in this population are prevalent; 64 percent of jail inmates, 54 percent of state prisoners, and 45 percent of federal prisoners report mental health concerns (The National Research

Council 2014). Many have experienced trauma, either before incarceration or while incarcerated, and resultant conditions such as post-traumatic stress disorder (PTSD) or anxiety may not be diagnosed. Yet only a minority receive treatment for their conditions while incarcerated (Kim et al. 2015). In addition to the need for mental health services, many people with justice system involvement need employment services. Employment is associated with reduced risk of recidivism (Kolbeck et al. 2022; Skardhamar and Telle 2012). Yet the unemployment rate among formerly incarcerated people is more than five times the rate among the general population (Couloute and Kopf 2018; The White House 2022).

IPS represents a promising model for helping this population to obtain jobs and succeed in the community. Studies have documented IPS' effectiveness for subgroups of people with severe mental illness (including those with justice involvement), but evidence of the effectiveness for adults with justice involvement who might not have severe mental illness is suggestive but limited, as it is from a small number of studies that had small sample sizes (Bond et al. 2015; Poremski et al. 2017; Doleac 2018).

IPS-AJI—a program that applies the IPS model to adults with recent justice involvement—was included in the NextGen Project for three reasons. First, it serves a population that may have diagnosed or undiagnosed mental health issues and hence may be at risk of needing SSI in the future if their mental health issues are unaddressed. Second, it uses an evidence-based model that could improve employment and reduce the need for SSI and other benefits. Third, although past studies have shown promising findings, no large-scale evaluation of IPS offered to adults re-entering the community from the justice system exists.

Overview of program and its logic model

This section provides a brief overview of IPS-AJI, what it is expected to change in the lives of the adults it serves, and how it is expected to change those lives (Exhibit B.1).

Participant referral sources

To recruit adults with justice involvement into IPS-AJI, the participating mental health centers partner with justice system organizations including jails, mental health or drug courts (which supervise people receiving alternative sentences that avoid imprisonment), and departments of probation and parole. Some centers also partner with homeless shelters as well as supportive or transitional housing programs to which justice system organizations make referrals. Finally, most centers also recruit study participants from their own mental health services caseloads.

Eligibility criteria

To be eligible to participate in the NextGen evaluation of IPS-AJI, an individual must consent to participate and meet the following criteria:

1. Be age 18 or older.
2. Be eligible for mental health services at the host mental health center.
3. Have either (1) been released in the past year into the mental health center service area from a justice system facility following the completion of a sentence, a case dismissal, or a verdict of "not guilty" within the past year; or (2) been ordered to probation or to serve an alternative sentence by a homeless, mental health, or drug court in the mental health center service community within the past year.

4. Be currently unemployed or doing casual jobs with no expectation of ongoing paid work or guaranteed weekly work hours.
5. Report during the IPS-AJI recruitment process that they are interested in working in the local area and are not prevented from working by probation, parole, or other court-ordered rules.
6. Have not received SSI or SSDI benefits as an adult in the past year and are not awaiting a decision on an SSI or SSDI application.

Exhibit B.1. IPS-AJI logic model

Participant needs and program resources	Program activities and services	Short-term outcomes	Long-term outcomes
Participant needs • Income • Mental health services **Program resources** • IPS specialists with maximum caseloads of 20 • Mental health professionals • Other mental health center staff and executives • Mental health center facilities and other services • Career profile for ongoing vocational assessment and job search/educational planning • Training, implementation support and performance feedback from the IPS Employment Center and state IPS trainers, including annual fidelity reviews and a workshop on communicating to employers about justice involvement • Departments of vocational rehabilitation	• *Zero exclusion*: IPS is open to anyone who wants to work, irrespective of their challenges; no one is screened out of IPS. • *Focus on competitive jobs*: The goal of IPS is for participants to obtain jobs in the competitive labor market rather than in sheltered workshops or community or transitional work experience positions. • *Rapid job search*: Participants engage in job search immediately rather than in lengthy assessments, job clubs, or workshops. • *Systematic job development*: Staff who provide employment services build relationships with employers and conduct individualized job development on behalf of participants. • *Participant-directed services*: Participants' preferences guide their job search, job choice, program staff interactions with employers, and educational planning. • *Integrated services*: IPS provides integrated mental health and employment services through a team-based approach. • *Availability of benefits planning*: IPS staff refer interested participants for comprehensive, individualized benefits planning by a trained benefits specialist. • *Time-unlimited supports*: IPS provides individualized long-term supports to help participants succeed on the job and with job changes and career advancement.	• Participants engage in rapid job search activities • Participants obtain a competitive job that matches their interests, strengths, and preferences • Participants engage in mental health and/or substance abuse services • Participants avoid the need for SSI/SSDI	• Participants achieve improved economic stability through steady employment in a competitive job, increased earnings, improved job quality, and reduced reliance on SSI/SSDI • Participants experience improved life satisfaction, self-esteem, mental health/substance abuse, and fewer hospitalizations • Participants experience lower levels of involvement in the justice system

Community context

- Availability of jobs
- Availability of skilled staff to hire
- Availability of other employment services
- Availability and accessibility of transportation
- Availability and accessibility of mental health treatment
- Culture around mental health and attitudes toward seeking mental health care (e.g., stigma among target population and community)
- Criminal justice system policies, procedures, and requirements

Participant needs and resources

The target population for IPS-AJI has two critical needs. The first is income. In addition to having to meet their basic needs, many adults with recent justice system involvement have immediate financial obligations such as restitution, court fees, or fines. Many aspire to work, but their justice involvement presents a barrier to getting hired by certain employers or for certain jobs. Guidance on which employers and jobs to pursue and how to handle their justice involvement in the application process could increase their chances of obtaining Income through earnings. The second need is for mental health care. As noted above, mental health issues are prevalent among people with justice system involvement, yet only a minority receive treatment for their conditions while incarcerated.

The NextGen Project pays for the IPS-AJI services provided to NextGen study participants in the program group. The project uses the following resources to provide services:

- **Staff.** IPS-AJI services are provided by IPS specialists with caseloads of no more than 20 individuals. IPS specialists receive support from supervisors and mental health center executives, and they collaborate with mental health professionals and other mental health center staff, such as case managers and benefits coordinators.
- **Program materials.** A career profile, which documents someone's work experience, interests, needs, and goals, informs ongoing vocational assessment, job search, and education planning.
- **Physical space.** Mental health services are provided at the mental health center offices, while IPS-AJI employment services are typically provided in the community.
- **Training and technical assistance**. The IPS Employment Center and state IPS trainers provide implementation support and performance feedback, using a fidelity review manual.
- **Partner agencies**. IPS-AJI programs collaborate with state departments of vocational rehabilitation, which can fund work supports that IPS may not be able to fund—such as work clothes or car repairs to enable participants to travel to and from work—and provide additional job coaching and other services.

Activities conducted and services provided

The core of IPS-AJI is the integration of employment and mental health services. IPS specialists help participants find jobs in the competitive labor market by getting to know their interests, strengths, and preferences and by building relationships with employers who offer jobs that fit those interests, strengths, and preferences. Job searching begins immediately upon a participant's enrollment in IPS-AJI; no lengthy assessments or pre-vocational activities occur, and no one is screened out of the program because of work challenges. Concurrent with employment services, participants receive mental health services from counselors, therapists, and/or psychiatrists. The IPS specialists, who provide services in the community, and mental health professionals, who provide services at the mental health center or virtually, collaborate in pursuit of the participant's career goals. IPS specialists provide employment services to participants for as long as participants desire; follow-along services for participants who become employed may include supports to succeed on the job and help with job changes and career advancement. The IPS fidelity review manual provides additional details on how IPS programs are intended to operate.

Participants who express interest are referred for benefits counseling to understand how work may affect any public assistance they may receive and to get help connecting to public assistance

programs. Participants may also be referred to a state department of vocational rehabilitation for work supports and other organizations for services the mental health center does not provide.

Expected outcomes

IPS-AJI services are expected to lead to four overarching short-term outcomes for participants:

- **Participant engagement in rapid job search**. The relationships that IPS specialists develop with participants are intended to help participants identify immediate job and ultimate career goals and apply for jobs that suit their interests, strengths, and preferences.
- **Increase in employment**. The rapid job search and job development activities are designed to connect participants with employers offering competitive jobs that match their interests, strengths, and preferences.
- **Participant engagement in mental health services**. The integrated, team-based approach aims to increase participants' engagement in mental health services.
- **Avoidance of SSI and SSDI application**. Earnings from competitive jobs will ideally be enough to preclude the need for SSI or SSDI.

Longer term, IPS-AJI aims to increase economic stability. Economic stability may include sustained employment, increased earnings, improved job quality, and reduced need for SSI or SSDI. For people with mental health issues, competitive employment has been shown to have positive effects on self-esteem and life satisfaction, and to reduce mental health symptoms. Limited evidence suggests that it may also result in longer-term reductions in psychiatric hospitalization, outpatient treatment, and other mental health service receipt (Luciano et al. 2014). By increasing financial security, routine activity, and commitment to a purpose, competitive employment may also lead to a reduction in criminal justice involvement (Apel and Horney 2017; Crutchfield and Pitchford 1997; Skardhamar and Telle 2012; Uggen 1999).

Descriptive and cost evaluations

Exhibit B.2 summarizes the evaluation-specific decisions made about the data collection for the descriptive and cost evaluations of IPS-AJI:

Exhibit B.2. Data collection sources and methods for the IPS-AJI descriptive and cost evaluations

Data collection	Respondents	Mode	Timing
Staff and leadership survey	Two or three executives per mental health center (such as mental health center director and clinical director); staff who recruit and work with NextGen study participants (such as research coordinators, IPS specialists and supervisors, mental health professionals, case managers, and benefits coordinators)	Web	Fall 2022
Fidelity reviews	Staff providing IPS-AJI and other mental health center services, their supervisors, and mental health center executives at each mental health center	In-person	Fall 2022; Fall 2023 for COCMHC

Data collection	Respondents	Mode	Timing
In-depth staff interviews	Staff providing IPS-AJI and other mental health center services, their supervisors, and mental health center executives at each mental health center	Telephone	Winter 2023; Fall 2023 for COCMHC
Employer interviews	In each mental health center service area, one employer with which IPS-AJI staff have developed a relationship	Telephone	Spring 2023
In-depth interviews with participants	Up to 15 IPS-AJI participants from across the five mental health centers who had met with an IPS specialist at least once	In-person and telephone	Spring 2023
Service receipt	People who had been enrolled in the study for at least six months as of June 2023	Each mental health center's management information system	Summer 2023
Study participant demographics	People who had been enrolled in the study for at least six months as of June 2023	Web-based RAPTER	Summer 2023
Cost collection	Select administrators and staff at each mental health center	Excel workbooks	Spring 2023; Winter 2024 for COCMHC

Note: Data from HOPE are not included in the descriptive or cost studies because HOPE began providing services much later the other mental health centers and served many fewer participants.

For the cost study, we used a cost study workbook in Excel to collect data on the costs of implementing IPS-AJI. We sent the tailored workbook to several mental health center administrators and staff (such as the center director, chief financial officer, accounting operations manager, IPS supervisor, and IPS program manager) and held telephone conversations with them in spring and summer 2023 to further explain the data request. We focused on costs incurred from July 1, 2022, through June 30, 2023, for Transitions and CMI; January 1, 2023, through June 30, 2023, for Grand and Pee Dee; and July 1, 2023, through December 30, 2023, for COCMHC. Each period reflects costs during a steady state of operations (that is, excluding the costs of initial implementation) and aligns with the timing of the descriptive study. We chose a shorter period for Grand and Pee Dee to exclude the cost of their IPS-AJI expansion efforts.

Impact evaluation

The impact evaluation is tailored to the evaluation of IPS-AJI in three main ways: (1) the point at which random assignment occurred, (2) the services offered to comparison group members, and (3) the data collected from study participants at baseline and at follow-ups.

Random assignment procedures

Random assignment occurs after a person has been determined to be eligible for IPS-AJI (including having or obtaining a mental health diagnosis from the mental health center), has consented to participate in the evaluation, and has completed a baseline survey. The person completes these study enrollment activities in person with a research coordinator at the mental health center or another location in the community. Before beginning the enrollment activities, the research coordinator talks with the individual about what is involved in participating in IP-AJI. These discussions may occur while the person is still incarcerated.

Services offered to the comparison group

Comparison group members may receive any services provided by the mental health center other than the employment services IPS specialists offer. Immediately after random assignment, a research coordinator refers members of the comparison group to a mental health professional for mental health services and refers members of the program group to an IPS specialist, who will provide employment services and introduce participants to their mental health provider.

Data collection

The data collection for the IPS-AJI evaluation differs from the data collection for the evaluations of other NextGen programs in three ways: (1) collecting data on justice involvement, (2) the length of the follow-up period, and (3) tailoring of the baseline and follow-up surveys.

- **Justice involvement**. To supplement data on criminal justice involvement from follow-up surveys, we are exploring collecting arrest, conviction, and incarceration data from state agencies. We are looking to collect criminal justice data in seven states: the four states the IPS-AJI programs operate in and three bordering states. We will try to get a sample of these data to determine the value of continuing to collect them for the final impact analysis. If data are obtainable and complete, we will match criminal justice records to IPS-AJI study participants (program and comparison group members).
- **Follow-up period**. The follow-up survey will occur at six months after random assignment and then again at 18 months after random assignment. We opted to conduct the first follow-up survey after six instead of nine months to maximize the survey response rate. The target population is very mobile and thus difficult to locate; a shorter period between the baseline and first follow-up survey should reduce sample attrition. A six-month follow-up period is consistent with when we would expect to see short-term impacts because of the rapid nature of the services. We selected a second follow-up period of 18 months to keep the time between the first and second follow-ups to only a year, which will reduce concerns about survey recall errors and locating study participants.
- **Tailoring the baseline and follow-up survey data collection**. The baseline and follow-up surveys were tailored for the evaluation of IPS-AJI. The main differences are shown in Exhibit B.3.

Exhibit B.3. Tailoring of the baseline and follow-up surveys for IPS-AJI

Change	Baseline	Follow-up	Rationale
Added a question about whether respondent is currently expected to pay child support.	✓		Anticipates that most of the target population will be male and will include many noncustodial parents.
Replaced question about currently working for pay with questions about working for pay in the month and year before the respondent's last arrest.	✓		Recognizes that many study participants will complete the baseline survey immediately after release from jail or prison and thus will not be currently working for pay; more relevant is whether they were working before incarceration.

Change	Baseline	Follow-up	Rationale
Added questions about respondent's experience with the criminal justice system, including the number of convictions of crimes and felonies, current form of court-ordered supervision, type of crime charged with, and incarceration history.	✓	✓	Baseline data provide additional description of the target population, and follow-up data allow for measurement of impacts related to recidivism.
Added questions about drug and alcohol use and receipt of services for problems related to drug or alcohol use.		✓	Substance and alcohol use are suspected to be high for the IPS-AJI population and may be affected by the IPS-AJI intervention.

References

Apel, R., and J. Horney. "How and Why Does Work Matter? Employment Conditions, Routine Activities, and Crime Among Adult Male Offenders." *Criminology,* vol. 55, no. 2, 2017, pp. 307–343.

Bond, G.. "Evidence for the Effectiveness of Individual Placement and Support Model of Supported Employment." IPS Employment Center 2022. https://docs.google.com/presentation/d/1RFqFrzidP_EwUEb_tgZ57LUJGpodbCM-7oQIQDda1P8/edit#slide=id.p72

Bond, G.R., S.J. Kim, D.R. Becker, S.J. Swanson, R.E. Drake, I.M. Krzos, V.V. Fraser, S. O'Neill, and R.L. Frounfelker. "A Controlled Trial of Supported Employment for People with Severe Mental Illness and Justice Involvement." *Psychiatric Services*, vol. 66, no. 10, 2015.

Couloute, L., and D. Kopf. "Out of Prison & Out of Work: Unemployment Among Formerly Incarcerated People." Prison Policy Initiative, July 2018. https://www.prisonpolicy.org/reports/outofwork.html.

Crutchfield, R.D., and S.R. Pitchford. "Work and Crime: The Effects of Labor Stratification." *Social Forces,* vol. 76*,* no. 1, 1997, pp. 93–118. https://doi.org/10.2307/2580319

Doleac, J. "Strategies to Productively Reincorporate the Former-Incarcerated into Communities: A Review of the Literature." Bonn, Germany: IZA Institute of Labor Economics, 2018.

Drake, R.E., D.R. Becker, and G.R. Bond. "Growth and Sustainment of Individual Placement and Support." Psychiatric Services, vol. 71, no. 10, 2020.

Frederick, D.E., and J.T. VanderWeele. "Supported Employment: Meta-analysis and Review of Randomized Controlled Trials of Individual Placement and Support." *PLoS One,* vol. 14, no. 2, 2019, p. e0212208. doi: 10.1371/journal.pone.0212208. PMID: 30785954; PMCID: PMC6382127.

KiDeuk, K., M. Becker-Cohen, and M. Serakos. "The Processing and Treatment of Mentally Ill Persons in the Criminal Justice System: A Scan of Practice and Background Analysis." Washington, DC: Urban Institute, March 2015.

Kolbeck, S.G., P.E. Bellair, and S. Lopez. "Race, Work History, and the Employment Recidivism Relationship." *Criminology,* vol. 60, no. 4, 2022, pp. 637–666.

Luciano, A., G.R. Bond, and R.E. Drake. "Does Employment Alter the Course and Outcome of Schizophrenia and Other Severe Mental Illnesses? A Systematic Review of Longitudinal Research." *Schizophrenia Research,* vol. 159, nos. 2–3, 2014, pp. 312–321. doi: 10.1016/j.schres.2014.09.010. Epub 2014 Sep 30. PMID: 25278105.

National Research Council. *The Growth of Incarceration in the United States: Exploring Causes and Consequences.* Committee on Causes and Consequences of High Rates of Incarceration, J. Travis, B. Western, and S. Redburn, Editors. Committee on Law and Justice, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press, 2014.

Poremski, D., D. Rabouin, and E. Latimer. "A Randomised Controlled Trial of Evidence Based Supported Employment for People Who have Recently been Homeless and have a Mental Illness." *Administration and Policy in Mental Health and Mental Health Services Research,* vol. 44, 2017, pp. 217–224.

Skardhamar, T., and K. Telle. "Post-Release Employment and Recidivism in Norway." *Journal of Quantitative Criminology,* vol. 28*,* no. 4, 2012, pp. 629–649. https://doi.org/10.1007/s10940-012-9166-x

The White House. "Expanding Economic Opportunity for Formerly Incarcerated Persons." Blog post. May 9, 2022. https://www.whitehouse.gov/cea/written-materials/2022/05/09/expanding-economic-opportunity-for-formerly-incarcerated-persons/.

Uggen, C. "Ex-offenders and the Conformist Alternative: A Job Quality Model of Work and Crime." *Social Problems,* vol. 46*,* no. 1, 1999, pp. 127–151. https://doi.org/10.1525/sp.1999.46.1.03x0245k

Appendix C: Further information about the design of the evaluation of Philly WINs

Under the NextGen Project, Community Integrated Services (CIS), a nonprofit serving Greater Philadelphia, provides employment services to adults with low incomes and disabilities who have sought support from the local workforce system. CIS provides the services through the Philadelphia Workforce Inclusion Networks (Philly WINs) program, which helps participants obtain substantial long-term employment while providing employers access to an untapped pipeline of talent.

This appendix provides information about why Philly WINs was selected for evaluation as well as the program and its logic model. Supplementing the discussion in Chapters 3-4 of the report, it also discusses how the descriptive, cost, and impact evaluations were tailored for the Philly WINs evaluation.

Rationale for evaluation

CIS has been providing employment services and opportunities to individuals with disabilities since 1991. Historically, CIS has collaborated with the Transition Pathways unit at A.J. Drexel Autism Institute (Transition Pathways) and other organizations to help youth with disabilities achieve a successful transition to adulthood. Just before the start of the COVID-19 pandemic, CIS, Transition Pathways, and an external consultant developed the Business Inclusion Center (BIC) model, through which they help companies implement best practices in employing individuals with disabilities in exchange for a commitment from them to hire and support workers with disabilities. Philly WINs, formally established in 2022 as part of the NextGen Project, is an outgrowth of this work.

In Philadelphia, American Job Centers, called PA CareerLink® Philadelphia centers (hereafter, CareerLinks), provide employment and training services for all types of job seekers. In addition to providing employment services funded by the Workforce Innovation and Opportunity Act (WIOA), Wagner-Peyser Act Employment Services, and other sources, CareerLinks serve people receiving Temporary Assistance for Needy Families (TANF) who county assistance office eligibility workers deem ready for work. Typically, eligibility workers consider most TANF recipients except those with medical exemptions from work requirements as ready for work, even if they have other complex challenges to employment such as unstable housing. While the CareerLinks serve people with a wide variety of challenges, they do not have specific programs or resources dedicated to the unique needs of people with disabilities. Expanding Philly WINs to the CareerLinks presents a promising approach for filling this gap.

Overview of program and its logic model

This section provides a brief overview of Philly WINs as designed, what it is expected to change in the lives of the adults it serves, and how it is expected to change those lives (Exhibit C.1).

Exhibit C.1. Philly WINs logic model

Participant needs and program resources	Program activities and services	Short-term outcomes	Long-term outcomes
Participant needs • Preparation for and access to quality jobs in environments supportive of people with disabilities • Support for reducing logistical barriers to employment (for example, child care and transportation) **Program resources** • CIS and Transition Pathways leaders, staff, and facilities • CareerLink workforce services and resources	**For participants** • *Vocational assessments.* Using standardized tools, staff assess participants on the skills and behaviors required for occupations in computer technology, business marketing, construction, manufacturing processes and production, and consumer services. • *Job readiness workshops.* Staff conduct three, two-hour workshops on (1) interviewing, (2) goal setting, and (3) time management. • *Individualized job search assistance.* Staff engage participants in job readiness activities (such as, creating a resume, identifying potential job leads, or practicing for a job interview); connect them to job openings; and support them with the job application and onboarding process. • *Access to employers in the CIS network.* Staff build relationships with BIC and other employers that can facilitate job placements for participants. • *Follow-along job support.* Staff support participants who are struggling while employed, lose their job, or need help finding a better job. • *Resource assistance.* Staff help participants access supportive services such as child care and transportation and provide referrals to address other complex issues that are or may become barriers to employment. • *Incentives.* Participants can earn up to $300 in incentives for engaging in program activities. **For employers** • *Training and technical assistance.* Staff support select employers in making hiring and onboarding processes accessible and company culture inclusive. • *Hiring events.* Staff facilitate hiring events tailored to meet select employers' needs. • *Job matching.* Staff help employers assess and fill their hiring needs by recommending suitable candidates from among Philly WINs participants. • *Job retention support.* Staff provide job coaching for employees who are Philly WINs participants to improve their performance.	• Participant engagement in job readiness and job search services – Assessments – Workshops – Job development meetings • Increase in participant employment – Job applications submitted – Participants hired – Hours working per week • Participant employment in quality jobs – Jobs at BIC employers – Jobs paying at least $15 per hour – Jobs offering full-time hours – Jobs with flexible hours – Jobs offering opportunities for advancement	• Participants maintain employment and advance in their careers • Participants achieve improved economic stability through increased earnings and reduced need for public benefits such as TANF or SSI/SSDI • Participants experience improved well-being with respect to life satisfaction, self-esteem, and mental health
Community context • Labor market conditions • TANF and WIOA policies	• Availability of and culture around transportation, child care, disability, and other employment services		

Participant referral sources

The study will enroll 1,000 individuals. The four CareerLinks are the primary referral source. Philadelphia's Department of Behavioral Health and other CIS partners also provide some referrals. Individuals who hear about the study in the community can self-refer.

Eligibility criteria

To be eligible for Philly WINs, individuals must meet the following criteria:

- Be 18 years old or older.
- Be able to work in the U.S. legally.
- Have a mental, emotional, or physical challenge to work that either is self-disclosed or identified by staff in the CareerLink who are already working with the individual. Philly WINs considers these conditions disabilities.
- Have low income. For the study, low income is defined as (1) receiving TANF, Supplemental Nutrition Assistance Program benefits, or refugee assistance or (2) having a household income below 200 percent of the federal poverty level, adjusted for household size.
- Speak English, Spanish, or use American Sign Language.

Participant needs and resources

The target population for the evaluation of Philly WINs is diverse and includes dislocated workers, individuals re-entering the community from the justice system, TANF recipients, and other adults with low incomes. By virtue of their visit to a CareerLink, all need support finding a job. However, because eligible participants are those with a mental, emotional, or physical challenge to work, they have a specific need for jobs that are accessible to people with disabilities and for employers that are supportive of people with disabilities. They also need support to reduce logistical barriers to work, such as transportation or legal assistance and, for TANF recipients in particular, child care assistance.

The NextGen Project pays for Philly WINs services that are provided to NextGen study participants in the program group. The project uses the following resources to provide services:

- **Staff.** Philly WINs services are provided by a variety of staff, including intake liaisons, vocational facilitators, job developers, and resource specialists. Philly WINs participants also have access to services provided by staff at the CareerLinks, including workforce advisors for those in TANF or WIOA and universal team members for others. Philly WINs staff are supported by program managers at CIS and Transition Pathways who have expertise in the best employment practices for people with disabilities. CareerLink staff are supported by program managers at Philadelphia Works, the organization that administers Philadelphia's Workforce Development Board and funds the CareerLinks.
- **Physical space.** Initial engagement of Philly WINs program group participants occurs during intake interviews, which take place primarily in one of the four CareerLinks. Assessments, workshops, and individual meetings between program staff and participants take place in the Northwest CareerLink or at the Transition Pathways office in West Philadelphia. No services are provided at CIS's office.

- **Funding for supportive services.** The TANF and WIOA programs can provide funding for work supports, such as child care or transportation, to Philly WINs participants. CIS also has funds to help cover some supportive services for Philly WINs program group members.

Activities conducted and services provided

Philly WINs begins with an intake session, a 60-minute guided discussion to get to know the participant, informally assess their well-being and general functioning, and gather preliminary information to share with the vocational facilitators (who conduct assessments and lead job readiness workshops) and job developers (who conduct job search assistance and follow-along job support). Participants then attend a two-hour Welcome to Philly WINs group session, where staff introduce them to the program and begin to build relationships with them. Participants complete their onboarding by engaging in a vocational assessment and attending three job readiness workshops. Throughout their participation in Philly WINs, participants meet individually with staff who provide job search assistance. They may attend as many career coaching sessions as they like. In these sessions, held twice per week at each Philly WINs service site, participants can get answers to questions about their job search and receive peer support from fellow participants. On an ongoing basis, staff build relationships with employers that can facilitate job placements for participants. Program staff support participants through the job application, hiring, and onboarding process. Once participants are hired, staff continue supporting them so that they can succeed on the job or, if necessary, find a different job. Participants may receive up to $300 in incentive payments, which are designed to encourage them to engage in program activities.

CIS and Transition Pathways work intensively with select employers in the greater Philadelphia area that meet the BIC criteria. To meet the BIC criteria, an employer is expected to offer high-quality jobs, hire and support workers with disabilities, and create an inclusive work culture. Philly WINs considers high-quality jobs to be those that pay at least $15 per hour, provide flexible hours and full-time positions, and offer advancement opportunities. Staff conduct initial and ongoing assessments of these employers' needs and provide customized services to meet them. They provide training and technical assistance in making hiring and onboarding processes accessible and company culture inclusive. They also facilitate hiring events tailored to meet BIC employers' needs and to educate them about the potential benefits of hiring and supporting people with disabilities, such as the Disability Work Opportunity Tax Credit or accessing a pool of qualified job seekers. CIS and Transition Pathways help both BIC and other businesses assess and fill their hiring needs by recommending suitable candidates from among Philly WINs participants.

Expected outcomes

Philly WINs services are expected to lead to three overarching short-term outcomes for participants:

1. **Engagement in job readiness and job search services.** The relationships that staff develop with participants are intended to encourage them to engage in program activities; identify their immediate job and ultimate career goals; and apply for jobs that suit their interests, strengths, and preferences.
2. **Increased employment.** Rapid job search and job development activities are designed to connect participants with employers offering competitive jobs that match their interests, strengths, and preferences.
3. **Employment in quality jobs.** The program's relationships with employers are intended to lead to jobs that enable participants to support themselves and their families. In addition to wages

and hours, other factors considered with respect to job quality include the degree of flexibility the job offers, the available accommodations for disabilities, and the opportunities for career advancement.

Longer term, Philly WINs aims to increase economic stability. Economic stability may include sustained employment; improved job quality; increased earnings; and less need for public benefits such as TANF, Supplemental Security Income (SSI), or Social Security Disability Insurance (SSDI). Employment has been found to have positive impacts on self-esteem, life satisfaction, and mental health for both the general population and people with disabilities (Drake and Wallach 2020; Kamerāde 2019; Modini 2016). Thus, an additional longer-term outcome is improvement in these measures of overall well-being.

Descriptive and cost evaluations

Exhibit C.2 summarizes the evaluation-specific decisions made about the data collection for the descriptive and cost evaluations of Philly WINs.

Exhibit C.2. Data collection sources and methods for the Philly WINs descriptive and cost evaluations

Data collection	Respondents	Mode	Timing
Staff and leadership survey	Six Philly WINs program leaders and 16 Philly WINs staff	Web	Summer 2023
In-depth interviews with staff	Interviews with CIS, Transition Pathways, and CareerLink center leaders and staff and with human resources staff at employer partners	In person	Spring 2023
Employer observations	Observations of job fair organized by CIS and discussions with other employers participating in the fair	In person	Spring 2023
In-depth interviews with participants	11 Philly WINs participants who received at least one hour of service from Philly WINs	Telephone	Spring 2023
Service receipt	People enrolled in the study for at least four months, as of August 31, 2023	CIS's management information system	Fall 2023
Study participant demographics	People enrolled in the study for at least four months, as of August 31, 2023	Web-based RAPTER	Fall 2023
Cost collection	Select administrators and staff at CIS	Excel workbooks	Winter 2024

For the cost study, the NextGen team used a workbook in Excel to collect data on the costs of implementing Philly WINs. The team sent the tailored workbook to CIS and held telephone conversations with them in winter 2024 to further explain the data request. The team focused on costs incurred from July 1, 2023, through December 31, 2023, to reflect costs during a steady state of operations (that is, excluding the costs of initial implementation).

Impact evaluation

The impact evaluation is tailored to the evaluation of Philly WINs in three main ways: (1) the point at which random assignment occurs, (2) the services offered to comparison group members, and (3) the data collected from study participants at baseline and at follow-ups.

Random assignment procedures

Random assignment occurs after a person has been determined to be eligible for Philly WINs, has consented to participate in the evaluation, and has completed a baseline survey. The person completes these study enrollment activities in person or over the phone with a Philly WINs liaison, usually at a CareerLink.

Services offered to the comparison group

Comparison group members may not receive Philly WINs services but may receive any services provided by CareerLinks or other employment services offered in the community. Immediately after random assignment, a Philly WINs liaison refers members of the comparison group to a workforce advisor or other staff person at the CareerLink who can help them explore the resources and programs at the center that can best support their job search and employment goals.

Data collection

The data collection for the Philly WINs evaluation differs from the data collection for the evaluations of other NextGen programs in three ways: (1) collection of data on TANF receipt, (2) the length of the follow-up period, and (3) tailoring of the baseline and follow-up surveys.

1. **TANF receipt.** To supplement data on TANF receipt from follow-up surveys, the NextGen team plans to collect data on TANF receipt from the Pennsylvania Department of Human Services. If the data are obtainable and complete, the team will match them to Philly WINs study participants (program and comparison group members) for analysis.
2. **Follow-up period.** The follow-up survey will occur at nine months after random assignment and then again at 21 months after random assignment. The NextGen team opted to conduct the first follow-up survey after nine months based on the length of time CIS typically works with participants. The team selected a second follow-up period of 21 months to keep the time between the first and second follow-ups to a year, which will reduce concerns about survey recall errors and locating study participants.
3. **Tailoring the baseline and follow-up survey data collection.** The baseline survey was tailored for the evaluation of Philly WINs by adding a question about whether the respondent is currently expected to pay child support, in anticipation that the target population would include many noncustodial parents. No customization was made to the follow-up survey.

References

Drake, R.E., and M.A. Wallach. "Employment Is a Critical Mental Health Intervention." *Epidemiology and Psychiatric Sciences,* vol. 29, November 5, 2020, e178. https://doi.org/10.1017/S2045796020000906

Kamerāde, D., S. Wang, B. Burchell, S.U. Balderson, and A. Coutts. "A Shorter Working Week for Everyone: How Much Paid Work Is Needed for Mental Health and Well-Being?" *Social Science & Medicine,* vol. 241, November 2019, 112353. https://doi.org/10.1016/j.socscimed.2019.06.006

Modini, M., S. Joyce, A. Mykletun, H. Christensen, R.A. Bryant, P.B. Mitchell, and S.B. Harvey. "The Mental Health Benefits of Employment: Results of a Systematic Meta-Review." *Australasian Psychiatry,* vol. 24, no. 4, August 2016, pp. 331–336. https://doi.org/10.1177/1039856215618523

Appendix D: Further information about the design of the evaluation of Western Mass MOMS℠

Western Mass Mental Health Outreach for MotherS Partnership℠ (Western Mass MOMS) serves adult caregivers who identify as women or nonbinary, have low incomes, and exhibit depressive symptoms. The program is based on the MOMS Partnership® model—a program designed to reduce depressive symptoms, improve social connections, and promote economic well-being among mothers. Western Mass MOMS is in Hampden County, Massachusetts, which includes Springfield and Holyoke.

The core of the MOMS Partnership model is a series of eight 90-minute weekly classes based on cognitive behavioral therapy (CBT) principles. The classes are designed to help participants manage their stress, better communicate with others, connect with their children, and work toward their goals. The class is facilitated by a clinician and a staff member who is a mother or caregiver from the community with lived experiences similar to those of the participants. The program also connects participants to needed supports and offers financial incentives for attending classes.

To tailor the program for the NextGen Project, we worked with the MOMS Partnership model developers at Elevate in the Yale School of Public Health and Viability Inc., the service provider, to enhance the program by offering employment services to Western Mass MOMS program participants. Employment services are offered both in groups—called Moving Forward groups—and, if needed, in one-on-one meetings with an employment specialist. The groups and the one-on-one meetings cover activities such as networking, resume development, job search, interviewing, or issues that arise on the job.

Western Mass MOMS is administered by Viability Inc. in partnership with the Massachusetts Department of Transitional Assistance (DTA). Viability is a nonprofit organization that provides services to individuals with disabilities as well as recipients of Temporary Assistance for Needy Families (TANF). DTA provides some referrals to Western Mass MOMS.

Technical assistance on the implementation of the program is provided by Elevate and The Adjacent Possible.[9]

This appendix provides information about why Western Mass MOMS was selected for evaluation and the program and its logic model. Supplementing the discussion in Chapters 3–4 of the report, it then discusses the formative evaluation that occurred and how the descriptive, cost, and impact evaluations are tailored for the Western Mass MOMS evaluation.

Rationale for evaluation

The MOMS Partnership model was developed as an evidence-based program to address concerns about elevated anxiety and depression among mothers with low incomes. It was developed by Dr. Megan Smith, an associate professor in the Yale School of Medicine and the Department of Social and Behavioral Sciences in the Yale School of Public Health.

[9] The Adjacent Possible, led by Dr. Michelle Derr, helped design the Western Mass MOMS evaluation and provided technical assistance about program implementation. Dr. Derr is a co-principal investigator for the NextGen Project. To avoid the appearance of a conflict of interest, staff from the Adjacent Possible did not collect or analyze data for the descriptive study, nor will they do so for the impact study.

MOMS Partnership has operated in New Haven, Connecticut, since 2011 and has served over 3,000 women. MOMS Partnership has also been implemented as part of the TANF program in Washington, DC, since April 2019 and in Vermont since February 2020. Additionally, since March 2020, MOMS Partnership has also served mothers with criminal justice involvement in Kentucky. These programs are funded primarily by the local health and human services or justice agencies.

The MOMS Partnership model is expected to reduce anxiety and depressive symptoms among its participants by offering a course based on CBT in a group setting. Many studies have found strong support for the efficacy of CBT for a wide range of psychological disorders, especially anxiety disorders (Hoffman et al. 2012). Providing the services in group settings addresses the social isolation that is common among mothers with low incomes and is associated with higher risks of depression (Plesko et al. 2021).

Participants' improved mental health is expected to lead to improved employment stability. Improved mental health can lead to increased employment through increased participation in education and training, increased likelihood of finding a job, and improved job retention (Banerjee et al. 2017; Morgenstern et al. 2009; Conrad et al. 1998). The enhanced employment services offered by Western Mass MOMS may further improve economic outcomes.

Western Mass MOMS—a demonstration program based on the MOMS Partnership model—was included in the NextGen Project for three reasons. First, it serves a population that experiences mental health issues and hence may need financial assistance, such as SSI, in the future if mental health issues are unaddressed. Second, some evidence suggests that the MOMS Partnership model could improve employment and economic independence, thus reducing the need for SSI and other benefits. Third, while past studies have shown promising findings, the MOMS Partnership model has yet to be rigorously evaluated.

Overview of program and its logic model

This section provides a brief overview of Western Mass MOMS, what it is expected to change in the lives of the caregivers it serves, and how it is expected to change those lives (Exhibit D.1).

Exhibit D.1. Western Mass MOMS logic model

Participant needs and resources	Program activities and services	Short-term outcomes	Long-term outcomes
Participant needs • Social support • Mental health services • Child care • Transportation • Financial support **Resources** • Relationship with DTA for referrals • Support provided by Yale University Elevate Policy Lab and The Adjacent Possible • Funding	• Staff training and ongoing technical assistance to train and support staff in delivering the MOMS Partnership model • Orientation • Stress Management Course, 8 classes – Curriculum based on cognitive behavioral therapy – Virtual and in-person courses – Three to 10 participants per class; cohort based – Co-led by a clinician and a Community Mental Health Ambassador (CMHA) – Incentives for participation in Stress Management Course and enrollment – Child supervision while participating • Employment services (including Moving Forward groups, one-on-one services, or services from other sources) • Contact with CMHAs between classes (on an as-needed basis) • Clinical supervision for direct service providers • Connections to community resources such as clothing, financial literacy education, toys for children, and emergency assistance • Community presence that builds rapport for the program and relationships with other social service providers	• Decrease in depressive and anxiety symptoms • Improvements in social support and trust • Improvements in self-regulation skills • Improvements in self-efficacy • Improvements in parenting skills	• Sustained decrease in depressive and anxiety symptoms • Sustained improvements in social support and trust • Sustained improvements in self-regulation skills • Sustained improvements in self-efficacy • Sustained improvements in parenting skills • Improved labor market outcomes: job placement, hours worked, retention, quality of work, earnings • Reduced use of public assistance • Reduced material hardships such as lack of food, not being able to pay for necessary medical services or medicines, unstable housing, and not being able to pay bills on time • Improved child outcomes[a]

Community context

- Availability of jobs
- Availability of skilled staff to hire
- Availability of other employment services
- Availability and accessibility of transportation
- Availability of and access to mental health treatment
- Culture around mental health service and community attitudes to seeking mental health care (e.g., stigma [self and community])

[a] Although Western Mass MOMS is predicted to improve child outcomes in the long term, child outcomes are not being measured as part of the NextGen project.

Participant referral sources

Western Mass MOMS participants are recruited from multiple sources. DTA caseworkers serving TANF and Supplemental Nutrition Assistance Program (SNAP) participants refer program participants to Western Mass MOMS. Western Mass MOMS staff also reach out directly to participants at DTA program orientations and DTA reaches out directly to participants about the program via text messages. Other referral sources include local mental health organizations, housing authorities, and public schools.

In addition, Western Mass MOMS conducts outreach directly to potential participants via attending community events (such as trunk-or-treat events at Halloween) and by posting flyers at places attended by mothers such as public housing, WIC offices, and medical offices.

Eligibility criteria

Western Mass MOMS aims to serve mothers and caregivers who exhibit symptoms of mental health conditions. Marketing materials state that the program is for caregivers who are "feeling stressed." Referral sources are told to refer caregivers who are showing depressive symptoms.

To be eligible to participate in the NextGen evaluation of Western Mass MOMS, the caregiver must consent to participate in the evaluation and:

1. Identify as a woman or nonbinary.
2. Be age 18 or older.
3. Be pregnant or caring for at least one child who is younger than 18 (the caregiver does not need to be a mother, but needs to be the primary caregiver).
4. Not be receiving Supplemental Security Income (SSI) or Social Security Disability Insurance (SSDI) or waiting for a decision on an SSI or SSDI application.
5. Be receiving TANF, SNAP, Medicaid, or WIC, or have a self-reported monthly household income before taxes less than 200 percent of the federal poverty level.

Eligibility requirements for the MOMS Partnership model, as designed by Elevate, include screening applicants for risk of clinical depression as shown by their responses to questions on the Center for Epidemiologic Studies Depression Scale (CESD-R). The program implementers chose not to use the CESD-R as an eligibility criterion for Western Mass MOMS.

The program also screens for suicidal ideation and psychosis. A clinician determines on a case-by-case basis whether the applicant is suitable for participation in Western Mass MOMS if they exhibit suicidal ideation or psychosis.

Participant needs and resources

In addition to their mental health needs, participants often are socially isolated and face material hardships. They are all primary caregivers and need child care to engage in services and employment. Many also need assistance with transportation to get to the in-person classes and for employment.

The NextGen Project pays for the Western Mass MOMS services provided to NextGen study participants in the program group. DTA, Elevate, and The Adjacent Possible also provide support, as described above.

Activities conducted and services provided

The core of Western Mass MOMS is the Stress Management Course. Guided by a curriculum and delivered using instructor and participant manuals, this course is based on CBT principles. It is co-led by a clinician (a mental health professional with an advanced degree and clinical license in social work, psychology, counseling, mental health therapy, or a related field) and a Community Mental Health Ambassador (CMHA). CMHAs are mothers or caregivers from the community who have lived experience similar to the experiences of program participants. The course is offered virtually or in person at community locations. It includes eight classes, offered weekly for eight weeks. There are three to 10 participants per course, and it is cohort based—after the second class, new participants cannot join. The course is offered in English and Spanish. Attendees are given a $25 gift card for attending each class. Child supervision is provided for caregivers attending in-person classes.

Participants are also offered the opportunity to receive employment services through additional group meetings or one-on-one. Moving Forward groups are optional group meetings that Western Mass MOMS participants can attend weekly to identify and make progress on their stress management and career goals. Groups were designed to be facilitated by an employment specialist, last 60 minutes, and be attended in person or via Zoom on a drop-in basis. Unlike the Stress Management Course, participants do not receive an incentive to attend the Moving Forward groups and child supervision is not offered. Participants who want more assistance with employment issues can meet one-on-one—in person or virtually—with an employment specialist or a job developer, as needed.

After participants enroll in the study, they are signed up for an in-person or virtual program orientation. The orientation may be one-on-one or in a group. At the orientation, participants hear about the services offered and the expectations for participating in the program.

CMHAs also contact participants between classes as needed to provide homework help, follow up with any issues raised during class, and support participants' class attendance. Viability has created a list of potential community resources that CMHAs use to provide referrals.

Yale University's Elevate Policy Lab train the Western Mass MOMS staff on the Stress Management Course curriculum and provide ongoing technical assistance to staff. A clinical supervisor also offers weekly supervision meetings with staff to ensure that the clinician and CMHA facilitate the Stress Management Course effectively.

Expected outcomes

Western Mass MOMS services are expected to lead to five overarching, short-term outcomes for participants. These will largely be measured by using participants' responses to a six-month follow-up survey.

- **Decrease in depressive symptoms**. The curriculum is designed to provide participants with skills and approaches to improve mental health. The social support from the group may also improve mental health.
- **Improvements in social support and trust**. The group sessions are designed to improve social support and trust. Social support can include practical help, material support, or emotional support provided by others (American Psychological Association). Social trust is a belief in the honesty, integrity, and reliability of others in the community (Taylor et al. 2007).

- **Improvements in self-regulation skills**. Self-regulation skills are the skills needed to finish tasks, stay organized, and control emotions (Cavadel et al. 2018). They are sometimes called soft skills, social and emotional skills, life management skills, executive skills, or executive functioning skills. Much of the curriculum involves participants practicing these skills.
- **Improvements in self-efficacy**. All aspects of the program are designed to improve participants' views of their own ability to accomplish their goals.
- **Improvements in parenting skills**. The communication and other self-regulation skills taught in the Stress Management Course are expected to also improve the participants' parenting skills.

Longer term (after six months), Western Mass MOMS aims for these short-term beneficial impacts to be sustained and for benefits in economic stability to emerge. Benefits in economic stability may include increased participation in education or training and/or improved labor market outcomes. Improved labor market outcomes could include a greater likelihood of finding a job, working more hours, keeping or advancing in a job, gaining a higher-quality job (such as one that offers fringe benefits or a higher wage), and earning more. With higher earnings, the participants are less likely to need public assistance such as TANF, SNAP, and SSI. Higher earnings, more referrals to economic supports, and better use of resources available may reduce the material hardships participants face—such as lack of food, being unable to pay for necessary medical services or drugs, unstable housing, and being unable to pay bills on time. We will measure these outcomes through participants' responses to an 18-month follow-up survey.

Formative evaluation

Towards the beginning of the study, we conducted a formative evaluation of Western Mass MOMS to ensure that it was a high quality implementation of the MOMS Partnership. We conducted it on four program components: (1) the orientation; (2) the Stress Management Course; (3) Moving Forward groups; and (4) outreach, recruitment, and enrollment into the study. Some of the formative evaluation occurred with pilot groups of participants who received program services but did not participate in the impact evaluation. Other formative evaluation occurred during the first months of the evaluation with study participants. We collected data from participants and referral agency staff via brief surveys. In addition, we held focus groups with program staff implementing each program component and staff referring potential participants to the program.

Descriptive and cost evaluations

Exhibit D.2 summarizes the evaluation-specific decisions made about the data collection for the descriptive and cost evaluations of Western Mass MOMS.

Exhibit D.2. Data collection sources and methods for the Western Mass MOMS descriptive and cost evaluations

Data collection	Respondents or source	Mode	Timing
Staff and leadership survey	16 Viability staff	Web	Late winter 2023
Observations	Six sessions, including orientation, Moving Forward groups, and Stress Management Course classes	Virtual	March and June 2023

Data collection	Respondents or source	Mode	Timing
In-depth staff interviews	17 Viability staff, including program managers, clinicians, CMHAs, intake specialists, employment specialists, job developer, child supervision staff, and the reflective supervisor	In-person and virtual	March 2023
Interviews with DTA staff	Six DTA staff	Virtual	March 2023
Interviews with Elevate and MOMS Partnership designer	Two current and one former Elevate staff	Virtual	March 2023
Interviews with referral partners	Three referral partner staff	Virtual	March 2023
In-depth interviews with participants	17 program group participants who had attended at least one Stress Management Course class; Aimed to interview half that attended classes virtually and half in-person, and four Spanish-speaking participants	In-person and telephone	March and June 2023
Service receipt	Program group members only	Web-based RAPTER	March 2022– TBD
Cost collection	Viability staff	Telephone conversations and emails	Spring and summer 2023

For the cost study, we used a cost study workbook in Excel to collect data on the costs of implementing Western Mass MOMS. We sent the tailored workbook to Viability and held telephone conversations in spring and summer 2023 to further explain the data request. We focused on costs incurred from January to June 2023 when Viability was implementing the program at scale. This period excludes costs incurred during the initial implementation period before the program was in a steady state and aligns with the timing of the descriptive study.

Impact evaluation

We tailored the impact evaluation design in three main ways: (1) random assignment procedures, (2) the services offered to the comparison group members, and (3) the data collected at baseline and on outcomes.

Random assignment procedures

Random assignment occurs after a caregiver has been determined eligible for Western Mass MOMS, has consented to participate in the evaluation, and has completed a baseline survey. The caregiver completes these study enrollment activities over the telephone with a Viability staff member. Before beginning the enrollment activities, a Viability staff member talks with the participant about what is involved in participating in Western Mass MOMS. During the enrollment interview, program staff also confirm that the participant does not have suicide ideation or psychosis to the extent that it would make participation in the program inappropriate.

Services offered to the comparison group

Comparison group members have no further contact with Western Mass MOMS staff after study enrollment. At the enrollment interview, the Western Mass MOMS staff may refer members of both the program group and the comparison group to services in the community to address mental health, employment, or other needs. The program pays members of the program and comparison groups $25 for attending the enrollment interview.

Data collection

The data collection for Western Mass MOMS differed from data collection for the other programs in three ways: (1) collecting data on TANF and SNAP receipt, (2) the length of the follow-up period, and (3) tailoring of the baseline and follow-up surveys.

- **TANF and SNAP receipt**. DTA will provide data on the receipt of TANF and SNAP benefits by all members of the program and comparison groups for a year prior to study enrollment through 18 months after study enrollment.
- **Follow-up period**. The first follow-up survey will occur at six months after random assignment and the second follow-up survey will occur at 18 months after random assignment. Even after accounting for participants not beginning the Stress Management Course right after study enrollment, we expect that at six months after study enrollment most program participants will have completed the eight-week course. We selected a second follow-up period of 18 months to keep the time between the first and second follow-ups to only a year and hence reduce concerns of survey recall errors.
- **Tailoring the baseline and follow-up survey data collection**. The baseline and follow-up surveys were tailored for Western Mass MOMS. The main differences are shown in Exhibit D.3.

Exhibit D.3. Tailoring of the baseline and follow-up surveys for Western Mass MOMS

Change	Baseline	Follow-up surveys	Rationale
Added the Center for Epidemiologic Studies Depression Scale (CESD-R) in place of the K-6 scale	✓	✓	CESD-R is a depression screening tool. It has been used in other studies of MOMS Partnership.
Added a question about the number of people who respondent could talk to for help or advice	✓	✓	Provides additional information about respondent's social support.
Added two questions about social trust, asking respondent whether they think most people can be trusted and whether they trust people in their neighborhood	✓	✓	Provides information about respondent's social trust.
Added questions about goal setting		✓	Provides some information about self-regulation. Goal setting is an important element of self-regulation.
Added questions about participation in services provided in a group		✓	Provides information about whether the comparison group is also receiving services in a group.

Change	Baseline	Follow-up surveys	Rationale
Added Healthy Families Parenting Inventory, Parenting efficacy subscale		✓	Provides information about the respondent's perceptions about their parenting.

References

Banerjee, S., P. Chatteriji, and K. Lahiri. "Effects of Psychiatric Disorders on Labor Market Outcomes: A Latent Variable Approach Using Multiple Clinical Indicators." *Health Economics,* vol. 26, no. 2, 2017, pp. 184–205. doi: 10.1002/hec.3286.

Conrad, K.J., C.I. Hultman, A.R. Pope, J.S. Lyons, W.C. Baxter, A.N. Daghestani, J.P. Lisiecki, et al." Case Managed Residential Care for Homeless Addicted Veterans: Results of a True Experiment." *Medical Care,* vol. 36, 1998, pp. 40–53.

MOMS Partnership website. Homepage. n.d. Accessed April 3, 2020. https://medicine.yale.edu/childstudy/services/community-and-schools-programs/elevate/our-work/scaling/partnership/

Morgenstern, J., A. Hogue, and J. McKay. "Does Coordinated Care Management Improve Employment for Substance-Using Welfare Recipients?" *Journal of Studies on Alcohol and Drugs,* vol. 70, 2009, pp. 955–963.

Plesko, C.M., Z. Yu, K. Tobin, and D. Gross. "Social Connectedness Among Parents Raising Children in Low-Income Communities: An Integrative Review." *Research in Nursing & Health*, vol. 44, no. 6, 2021, pp 957–969.

Smith, M.V., L.S. Callinan, C.S. Posner, S.C. Holmes, and R. Ebling. "Improving Maternal Mental Health as a Pathway to Economic Mobility in the TANF System." *Psychiatric Services,* 2021.

Smith, M.V., L.S. Callinan, and M. Ciarlegio. "Can Treating Maternal Depression with Cognitive Behavioral Therapy Increase Employment Among Low-Income Mothers? A Pilot Study." *Journal of Health Care for the Poor and Underserved* (2021, under review).

Taylor, P., C. Funk, and A. Clark. "Americans and Social Trust: Who, Where and Why." Washington, DC: Pew Research Center, 2007. https://www.pewresearch.org/social-trends/2007/02/22/americans-and-social-trust-who-where-and-why.