Alternative Supporting Statement for Information Collections Designed for Research, Public Health Surveillance, and Program Evaluation Purposes

Next Generation of Enhanced Employment Strategies Project

OMB Information Collection Request
New Collection

Supporting Statement

Part B

March 2020

Submitted By:
Office of Planning, Research, and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

4th Floor, Mary E. Switzer Building
330 C Street, SW
Washington, D.C. 20201

Project Officer:
Hilary Bruck

Part B

B1. Objectives

Study Objectives

The Office of Planning, Research, and Evaluation (OPRE) within the Administration for Children and Families (ACF) at the U.S. Department of Health and Human Services (HHS) seeks approval for data collection activities conducted for the Next Generation of Enhanced Employment Strategies Project (NextGen Project). The objectives of this project are:

1. To identify and rigorously evaluate the effectiveness of about 10 innovative programs designed to promote employment and economic security among people with complex challenges to employment
2. To describe the operations, implementation successes and challenges, and lessons learned for each program
3. To estimate the costs of each studied program

As described in Supporting Statement A, the NextGen Project will use a two-phased approach for OMB approval of this ICR. This current request is for Phase 1, during which we will administer the informed consent form (Appendix A) and baseline survey (Instrument 1), and collect identifying and contact information (Instrument 2) for study participants. Under Phase 2, we will submit the remaining instruments, which we expect will require some revisions to tailor to each site selected for the evaluation. We are not seeking approval for Phase 2 instruments at this time but have included drafts of these instruments, burden estimates, and current details for initial review and informational purposes (Appendices F and H – O).

Generalizability of Results

The impact studies are intended to produce internally-valid estimates of the programs' causal impacts, not to promote statistical generalization to other sites or service populations. The descriptive and cost studies are intended to present internally-valid descriptions of the service population, implementation, and cost of the chosen programs, not to promote statistical generalization to other sites or service populations.

Appropriateness of Study Design and Methods for Planned Uses

The study's purposive selection of sites and its impact, descriptive, and costs studies are appropriate for the government's goal of identifying and rigorously evaluating innovative programs designed to promote employment and economic security among low-income people with complex challenges to employment. Resulting publications will enhance knowledge of the effectiveness of innovative employment strategies among federal, state, and local policymakers, practitioners, and other stakeholders and support peer learning.

- **Impact studies**. The project team will conduct a separate impact study for each program. The impact study will involve a randomized controlled trial (RCT), where participants eligible for participation in the program are randomly assigned to a treatment group that is offered the program or a control group that is not offered the program but can participate in other services normally available in the community. RCTs ensure that the

treatment and control groups are similar, on average, before the treatment group is offered the program; this means that any observed differences in outcomes between the groups can be attributed to the program rather than other factors. The results of the impact studies could be used to inform federal, state, and local policymakers about future funding of the tested programs; by program administrators and directors who might consider implementing the tested programs, or something like them, for their own programs; and program developers and technical assistance providers facilitating implementation of evidence-based practices. Therefore, it is important that these rigorous methods are used to assess effectiveness.

- **Descriptive studies.** The descriptive study for each program will describe the community, economic, and program context in which the program operates; the characteristics of the program model, including the target population, services offered, role of partners and employers, theory of change, and plans for sustainability and replication; and the implementation and cost drivers of the program, such as leadership, organizational culture and structure, staffing and staff development, and service delivery. This information will support interpretation of the impact findings and is critical in helping other programs replicate or refine the program for their own contexts.

- **Cost studies.** The cost study for each program will provide descriptive information about the amount, sources, and types of funding for the program, and estimate the average cost of the program per participant. In a benefit-cost analysis, the cost of the program per participant will be compared with an array of benefits, including potential gains in earnings and potential reductions in the receipt of alternative services and public assistance. This information is important for program funders and for practitioners considering implementing the programs.

As noted in Supporting Statement A, this information is not intended to be used as the principal basis for public policy decisions and is not expected to meet the threshold of influential or highly influential scientific information. One of the limitations of the study for its intended use (informing the design and adoption of future employment programs) is that the impact studies will produce an internally valid estimation of the impact of the program model and its implementation at the time of the study. Impacts are a function of the program model, its implementation, characteristics of the service population, and also of other similar services that are available in the studied community (sufficient treatment/control contrast). This context should be considered when interpreting whether and how findings will apply to program expansion; this will be stated in written products presenting study findings.

B2. Methods and Design

Target Population

The target population for the NextGen Project is low-income individuals with complex challenges to employment, including physical and mental health conditions, a criminal history, or limited work skills and experience. The project is working closely with the Social Security Administration (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 (SSI). Each selected program might focus on participants with one or more of these challenges.

The project will target a sample size of about 1,000 study participants for each program, with 500 each in the treatment and control groups. This will lead to a total sample size of about 10,000 participants across participating programs.

Sampling and Site Selection

The NextGen Project is currently identifying and assessing innovative programs on a rolling basis for inclusion in the study. The site selection approach is described in detail in two previous Generic ICR submissions, one for stakeholder engagement and one for site assessment, both of which received approval under the generic clearance for Formative Data Collections for ACF Research (OMB #0970-0356)[1]. In summary, the programs for the project will be selected to meet three general criteria:

1. The program addresses OPRE's, and in some cases SSA's, research interests.
2. The program is well implemented, or could be after some technical assistance.
3. It is feasible to rigorously evaluate the program using an experimental design, or could be after the program received evaluation technical assistance.

Sampling for impact studies. The sample frame for the impact study will be all people who are eligible for and interested in the program and consent to participate in the evaluation during the enrollment period. The project team will collect survey information from all study participants at three points: (1) at baseline, before random assignment occurs; (2) at about 6 to 12 months after random assignment via the first follow-up survey; and (3) at about 18 to 24 months after random assignment via a second follow-up survey. The project team will attempt to survey the universe of study participants and will also collect administrative data on earnings, benefit receipt, and potentially other outcomes on all study participants. The team anticipates obtaining information from administrative sources for all sample members.

Table B.1 reports program-level minimum detectable impacts on earnings outcomes for survey and administrative data. The target sample size for each study is 1,000 study participants, 500 each in the treatment and control groups. For the administrative data, the project team expects nearly 100 percent coverage; they expect about an 80 percent response rate for each of the participant follow-up surveys.

The minimum detectable impacts of a survey sample of 800 and an administrative data sample of 1,000 are about $549 and $492 in average quarterly earnings, respectively. This means that the study is powered to detect a difference of $549 or greater in quarterly earnings between the treatment and control groups as measured through the follow-up surveys. If the true difference in earnings between the groups is less than this, the study will likely not detect a statistically significant impact. 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. Evidence reviews, such as the What Works Clearinghouse, consider effect sizes of 0.25 standard

[1] Information collection activities for stakeholder engagement were approved on February 13, 2019 and activities related to site assessment were approved on June 4, 2019.

deviations or larger as substantively important (What Works Clearinghouse 2017), which the study will be able to detect. However, note that these are only estimates based on data from previous studies on similar populations.

Table B.1. Minimum detectable impacts per program on key outcomes for an RCT

Data source	Analysis sample size (treatment and control)	Quarterly earnings (impact, dollars)	Effect size
Follow-up surveys	800	549	0.18
Administrative data	1,000	492	0.16

Assumptions: individuals are randomly assigned; equal random assignment probabilities to treatment and control groups; $3,102 standard deviation of earnings; covariates explain 20 percent of the variation in the outcomes; response rate of 80 percent on the survey; two-tailed test, *p*-value of 0.05.

Sampling for descriptive studies. The descriptive studies will be based on three types of data collection, some of which involve purposeful sampling:

1) *Semi-structured discussions with program staff, leaders, and, if applicable, partners and employers.* Program staff and leaders will be selected purposively for discussions using organizational charts and information on each employee's role at the organization. Staff from partner organizations and employers, if applicable, will be selected based on their involvement with the program and its participants. Purposeful staff selection is appropriate because particular insights and information available from individuals will depend on their perspectives based on their role at the organization. The results of the descriptive study are not intended to generalize beyond the program being studied.
2) *Surveys of program staff and leaders.* The universe of all frontline staff and leaders at the selected programs will be asked to complete a web-based staff or leadership survey collecting information on their professional backgrounds, skills, experience, and perceptions of the program. This will provide a broader perspective on these topics than can be elicited through the interviews, and thus targeting the universe of staff and leaders is appropriate.
3) *In-depth interviews of program participants.* The project team will recruit approximately 20 treatment group members from each program to complete the interviews among treatment group members who have participated in the program. The team will select treatment group members who were randomly assigned at least six months before the interviews so that they will include study participants who have potentially participated in the program for six months. These interviews are to provide narrative, in-depth context and experiences of program participants.

Sampling for cost studies. Leaders from each program (or their designees) may submit their accounting records to the project team, who will use them to complete a standardized Excel-based workbook. Or, program staff who are familiar with the program's expenditure and accounting records may directly complete the workbook.

B3. Design of Data Collection Instruments

Development of Data Collection Instruments

Table B.2 lists the data collection instruments for the study and links them with the study's objectives. The data collection instruments were developed to capture essential data for the study's main research questions that are not readily available from administrative sources.

A description of how each question in the baseline survey, identifying and contact information, and follow-up surveys will be used in the analysis is provided in Appendices B, C, and D. Appendices B and D link each question to its objective in the analysis. These appendices also include references for items that were used in previous studies.

Phase 1 data collection instruments are final and are not expected to change for the purposes of the study, other than to drop questions from the baseline survey for specific types of programs. The program-specific questions are noted in Instrument 1: Baseline survey. Phase 2 instruments will be tailored for each selected program to reflect each program's theory of change.

Data for impact studies. The baseline and two follow-up survey instruments for the NextGen Project were developed by content experts at Mathematica and OPRE, and informed by reviewing instruments used in similar data collection efforts. Many questions are sourced from prior studies, such as the Parents and Children Together study (OMB #0970-0403), the Evaluation of Supplemental Nutrition Assistance Program (SNAP) Employment and Training Pilots (OMB #0584-0604), Evaluation of Employment Coaching for TANF and Related Populations (OMB #0970-0506), and the National Beneficiary Survey (OMB #0960-0800). Other items come from scales that have been frequently used in large-scale national surveys, such as the SF-12® Health Questionnaire to assess health status. Finally, the surveys were developed in coordination with the OPRE Building Evidence on Employment Strategies for Low-Income Families Project (BEES), with which the NextGen Project is coordinating closely (as described in Part A). Areas of overlap with the BEES instruments are described in the question-by-question justifications for the baseline and follow-up surveys (Appendices B and D).

The project team used industry best practices to reduce potential sources of measurement error. These practices include:

- Using validated items from previous surveys administered to similar populations to the extent possible.
- Including in the instruments automatically enforced skip patterns, built-in range checks, internal item consistency checks, and required answer fields.
- Pretesting the baseline and follow-up surveys with individuals similar to the populations served by the type of programs being assessed for inclusion in the NextGen Project. The project team timed the interviews and used cognitive interviewing and respondent and interviewer debriefings to assess respondents' understanding of the survey questions, identify improvements to the flow and structure of the instruments, and to ensure burden was low. The same question was not asked of more than 9 people. The surveys were updated based on the findings.

Data for descriptive studies. The discussion guides for program staff, employers, and partners, the surveys of program staff and leaders, and the in-depth participant interview guide were developed by content experts at Mathematica and OPRE. They were informed by reviewing

instruments used in similar data collection efforts. These efforts included the Evaluation of SNAP Employment and Training Pilots (OMB #0584-0604) and the Evaluation of Employment Coaching for TANF and Related Populations (OMB #0970-0506). The guides were also informed by a review of corresponding instruments submitted to OMB by BEES (OMB #0970-0537).

The project team pretested the staff and leadership surveys with staff and leaders, with similar background and work experience to those implementing programs being considered for inclusion in the NextGen Project. The same question was not asked of more than 9 people. As a result of the pretests, the surveys were updated for clarity, flow, and to reduce burden.

Data for cost studies. The Excel-based cost workbook was developed by Mathematica staff, who selected the cost elements based on cost-collection tools developed for the Evaluation of SNAP Employment and Training Pilots (OMB #0584-0604) and the Workforce Investment Act (WIA) Adult and Dislocated Worker Programs Gold Standard Evaluation (OMB #1205-0504). The project team does not intend to pretest the cost workbook since the team will provide training to program leaders (or their designees) on the cost study and how to complete the workbook, and a designated site liaison will work with programs to help them understand the request and complete the workbook accurately.

Table B.2. Crosswalk between the data collection instruments and the study's objectives

	Study objectives		
Instrument	Estimate the effectiveness of the program	Describe program operations and implementation	Estimate program costs
Instrument 1 (final): Baseline survey	All items	All items	n.a.
Instrument 2 (final): Identifying and contact information	All items (used to match to administrative outcome data and locate study participants for follow-up surveys)	n.a.	n.a.
Appendix F: Instrument 3 (draft): First follow-up survey	Items in Sections A, B, and C	Items in Section D	n.a.
Appendix H: Instrument 4 (draft): Second follow-up survey	Items in Sections A, B, and C	n.a.	n.a.
Appendix I: Instrument 5 (draft): Service receipt tracking	All service receipt items	All service receipt items	All service receipt items

Appendix J: Instrument 6 (draft): Staff characteristics survey	n.a.	All questions	n.a.
Appendix K: Instrument 7 (draft): Program leadership survey	n.a.	All questions	n.a.
Appendix L: Instrument 8 (draft): Semi-structured program discussion guide	n.a.	All questions	n.a.
Appendix M: Instrument 9 (draft): Semi-structured employer discussion guide	n.a.	All questions	n.a.
Appendix N: Instrument 10 (draft): In-depth participant interview guide	n.a.	All questions	n.a.
Appendix O: Instrument 11 (draft): Cost workbook	n.a.	n.a.	All items

Note: For Phase 1 of this ICR, clearance for Instruments 1 and 2 is requested. The remaining instruments will be submitted as Phase 2 of this ICR.

n.a. = not applicable.

B4. Collection of Data and Quality Control

Impact studies. The project team will collect the data for the impact evaluation via three surveys of study participants as well as administrative records.

- **Baseline survey (Instrument 1) and Identifying and contact information (Instrument 2).** In each program selected for the study, program staff will identify individuals eligible to participate in the program and administer the consent form (Appendix A) to the applicant. If the applicant consents to participate in the study, staff will enter the person's identifying and contact information into the Random Assignment, Participant Tracking, Enrollment and Reporting system, or RAPTER® (Instrument 2). Program staff will either administer the baseline survey (Instrument 1) to the program applicant or the applicant will self-administer the survey via the web. In some cases, if needed, study participants could call Mathematica's Survey Operations Center to complete the survey by phone with a trained telephone interviewer. After study participants have completed the baseline survey, program staff will ask for some contact information and enter it into RAPTER®. After this information is collected, RAPTER® will randomly assign each study applicant to the treatment or control group and notify program staff of the assignment. The program staff will notify the study participant of his or her assignment.

  To ensure quality and consistency in this data collection, the project team will:

  - Provide a written procedures manual to program staff who will enroll study participants.
  - Provide training to all program staff who will administer the survey. The training will cover administering consent, collecting identifying and contact information from participants, administering the baseline survey, notifying enrollees about the

result of random assignment, and handing them off to the correct post-assignment protocol (treatment or control). Additional trainings will be provided for new staff and if issues arise.

- Provide a designated liaison that the program staff can call to answer questions.
- Provide a hotline that the program staff can call if they cannot reach the designated liaison.

The project team will monitor for quality and consistency in the data collection by the program staff. They will regularly review the data entered into RAPTER® and the survey responses, looking for patterns of missing data and other data quality issues. They will work with programs to resolve them quickly.

- **First follow-up survey (Appendix F) and second follow-up survey (Appendix H).** The follow-up surveys will be available to all study participants to either self-administer via the web or complete using computer-assisted telephone interviewing (CATI).

  The project team will ensure quality and consistency in the data collected by the surveys by using a number of tactics, such as:

  - For self-administered web surveys: use clear and straightforward language; include predominantly closed-ended questions; include check boxes, drop-down menus, and response categories; include soft checks to prevent outlier entries; and ensure the layout is compatible with multiple browsers, tablets, and smartphones.
  - Recruit qualified interviewers to administer the survey by CATI.
  - Train the interviewers in interviewing techniques as well as the intent of each question in the survey.
  - Listen to about 10 percent of all CATI interviews to listen for inaccurate presentation of information on the study; errors in reading questions; biased probes; inappropriate use of feedback in responding to questions; and any other unacceptable interviewer behavior.
  - Examine data on the number of completed interviews, calls made, refusals, refusal conversions as well as time per call, and time per interview by interviewer. Supervisors will provide feedback to interviewers based on these data.
  - Debrief with groups of interviewers shortly after the start of a data collection to discuss the respondents' level of cooperation and ability to understand and answer the survey questions.
  - Examine frequencies and cross-tabulations on data collected on a regular basis to pinpoint any unexpected aspects of instrumentation, particularly in skip logic, valid value ranges, the operation of edits and consistency checks, and the recording of data for legitimately skipped items and "don't know" and refusal responses.
  - Examine frequencies and cross-tabulations on data collected, by mode of collection, to look for evidence of mode bias or large differences in responses between self-administered web surveys and interviewer-administered telephone interviews.

Descriptive studies

- **Service receipt tracking (Appendix I).** Program staff will use RAPTER® to record information about all treatment group members' participation in the program. If a program already collects data on service receipt through its own database, the project team will use the information the program already collects. To ensure quality and consistency in collection, the project team will train program staff involved with the evaluation on how to use RAPTER® to enter service receipt data with accuracy and in a timely fashion. The project team will monitor for quality and consistency in the staff entries into RAPTER® by reviewing what they are entering. Approximately twice per month, the team will check that program staff are entering data regularly and in as much detail as needed.

  If programs are using their own database to collect service receipt data, the project team will work closely with the program staff to understand what information is entered into the program's database, how staff enter information, and if any improvements are needed to ensure that the data will meet study needs and quality expectations. The project team will request a deidentified sample extract before the evaluation begins to make sure the program's information collection meets expectations. After the study begins, the project team will request data extracts regularly to ensure that data is received for all enrolled study participants. The project team will develop reports using the data to help monitor the data entry frequency and quality.

- **Staff characteristics survey (Appendix J) and Program leadership survey (Appendix K).** The project team will ask frontline staff and leaders to complete the appropriate survey via the web. To ensure quality and consistency in collection, the project team:

  - Designed the surveys to use clear and straightforward language; include predominantly closed-ended questions; include check boxes, drop-down menus, and response categories; and include program checks to prevent outlier entries.
  - Will regularly examine the data collected through the surveys, checking for indicators of potential quality issues such as blank open-ended responses or high item nonresponse rates.

- **Semi-structured program discussion guide (Appendix L) and semi-structured employer discussion guide (Appendix M).** The project team members will interview program staff, partner staff, and employers in person or by telephone. The project team will recruit program staff and leaders for discussions using organizational charts and information on each employee's role at the organization. The team will recruit staff from partner organizations and employers, if applicable, to offer perspectives based on their involvement with the program and its participants.

  To ensure quality and consistency in data collection, all interviewers will be trained in the study research questions, the research approach, the topics to be covered in the data collection, and techniques for data collection (including protecting privacy, preparing post-visit summaries, and ensuring data security). The training will also cover the content of the instruments to ensure full understanding of the questions and the collection of comparable, complete, and high quality data across the team. Any necessary refresher training will be provided. If respondents consent to being recorded, the interviewer will

audiorecord discussions with program administrators, supervisors, staff; and key partner staff, including employers. Task leaders for the descriptive study will periodically review completed interviews for quality and for missing information.

- **In-depth participant interview guide (Appendix N).** The project team will conduct in-person, one-on-one interviews with randomly selected study participants. The project team will contact participants to schedule interviews. If participants cannot be reached or refuse to participate in the interviews, the project team will replace that study participant with another randomly selected participant.

  To ensure quality and consistency in data collection, the interviewers will be trained in the study research questions, the research approach, the topics to be covered in the data collection, and techniques for data collection. Topics to be discussed include protecting privacy, using culturally appropriate and trauma-informed interviewing techniques, and ensuring data security. The training will also cover the content of the interview protocol to ensure full understanding of the questions and the collection of comparable, complete, and high quality data across the team. The project team will monitor for quality and consistency in the data collection. Any necessary refresher training will be provided. If respondents consent to being recorded, the interviewer will audiorecord discussions with participants.

Cost studies.

- **Cost workbook (Appendix O).** The project team will send an Excel-based workbook for collecting data on program costs to program leaders (or a designee) for each program. The workbook will record information on the expenditures associated with the program for a recent 12-month period.

  The data collection approach will include two steps. First, the project team will ask program leaders for their accounting records or financial reports and obtain as much information as possible from these records. Second, if additional information is needed after review of financial records, the project team will ask the programs to complete the workbook in part or in full, depending on the information required. One tab of the workbook is designed to collect information specifically from social enterprises; this tab will not be completed by programs that are not social enterprises.

  To ensure quality and consistency in collection, the project team will train the program leaders (or their designees) on the cost study and how to complete the workbook. A designated site liaison will work with programs to help them understand the request and complete the workbook accurately. The project team will monitor for quality and consistency in the data collection by thoroughly reviewing each completed workbook, checking for completeness and internal consistency. Interviewers may ask follow-up questions about the information entered into the workbook.

B5. Response Rates and Potential Nonresponse Bias

Response Rates

The project team will calculate conditional response rates as the number of completed surveys or other data collection instruments as a percentage of the number of people asked to complete the

survey or instrument. If any study enrollees become ineligible for the study after they have been randomly assigned, the project team will remove them from the denominator of the response rate calculation. This could happen if, for example, someone died during the course of the study.

Item response rates will be calculated as the number of people who complete an item as a percentage of the number of people who respond to any questions on the survey or other data collection instrument. The project team will exclude from the item response calculation any people who were not offered the question due to a survey skip pattern.

Impact studies

- **Baseline survey and Identifying and contact information.** Applicants eligible for study participation will only be enrolled in the study and randomly assigned if they complete the baseline survey and provide their identifying information as part of the intake process. Therefore, the project team anticipates that 100 percent of study participants will provide these data. The project team does not anticipate significant item nonresponse based on prior experience asking similar questions with similar populations. In a similar baseline survey that was used for the Evaluation of Employment Coaching for TANF and Related Populations (OMB #0970-0506), the item nonresponse was low; for example, the nonresponse to the employment status question was less than 3 percent.
- **Follow-up surveys.** The project team anticipates an 80 percent response rate on the follow-up surveys based on their experiences conducting follow-up surveys with similar populations and studies. The team will attempt to complete both first and second follow-up surveys with the entire sample. The project team has achieved similar response rates in other studies with hard-to-engage populations. For instance, in the evaluation of the Building Nebraska Families program (OMB #0970-0246), the team achieved an 87 percent response rate on the 18-month follow-up survey and an 83 percent response rate on the 30-month follow-up survey for TANF recipients who faced multiple challenges to employment. For the Personal Responsibility Education Program (PREP) evaluation (OMB #0970-0398), the project team achieved an 84 percent response rate on the 12-month follow-up survey and an 82 percent response rate for the 24-month follow-up survey for the Healthy Families San Angelo program, a home visitation program that seeks to engage a low-income population. For the Parents and Children Together follow-up surveys, the project team achieved an 88 percent response rate for the low-income mothers and fathers in the healthy marriage program study (OMB #0970-0403). The project team does not anticipate significant item nonresponse on the follow-up survey based on prior experience asking similar questions with similar populations, as described above.

  To maximize response rates on the surveys, the following techniques will be used, which were also employed in the aforementioned efforts:

- **Allow respondents to complete the survey in different ways.** Respondents will be able to complete the survey either online (using a computer, tablet, or smartphone) or by telephone.
- **Send reminder notifications.** In addition to notifying the study participant about the follow-up surveys during study intake, the project team will use a combination of letters,

emails, texts, and telephone calls to encourage people to participate (Appendix G) throughout data collection efforts.

- **Obtain accurate, up-to-date contact information.** The project team will collect detailed contact information during study intake and the follow-up surveys to aid in locating participants to complete the follow-up surveys. Before the start of the follow-up surveys, the project team will update participant contact information through online database searches and might also request updates from participants via text message or email.
- **Use intensive locating methods, as needed.** The project team will initially notify participants about the survey by mail and email and ask them to complete it via the web, though they will also be able to complete it via telephone at that time. After four weeks, the project team will attempt to contact the participants via telephone, so they can complete the survey via telephone. If the participants cannot be reached by telephone, the project team will contact the contacts identified by the participant during the baseline data collection, for help locating them. If the participants still cannot be located, the project team will conduct customized, individual searches for contact information using specialized databases. Finally, if study participants still cannot be located, trained field locators will go in person to the study participant's home and neighborhood. If they locate the study participant, the field locators will lend him or her a smartphone to complete the survey.
- **Offer tokens of appreciation.** As discussed in greater detail in Part A, Section A9, the study's proposed strategy for tokens of appreciation is designed to retain respondents in the longitudinal data collection and decrease the differential response rate between the treatment and control groups, and therefore reduce nonresponse bias on impact estimates.
- **Continuous quality improvement.** The study will collect data on each attempt to contact a respondent, including the mode, time, date, interviewer, and contact results. Examining these paradata will help identify the most effective calling times and interviewers. The project team will also use paradata to determine which methods of contact (letters, emails, texts, or telephone calls) are proving to be the most successful in this study, so that they can adjust the frequency and type of contacts to achieve high response rates.

Descriptive studies

- **Service receipt tracking.** Because programs are opting to participate in the study and will receive assistance from the project team to support their service receipt tracking, the team expects that program staff will enter data on all service receipt by members of the treatment group. The project team will monitor how programs enter program service receipt into RAPTER® and will encourage programs to keep the records up to date.
- **Staff and program leadership surveys.** Based on similar research projects, the project team expects program leaders will support staff's completion of the surveys, resulting in a high response rate among staff and leaders (around 90 percent). On the Evaluation of Employment Coaching for TANF and Related Populations (OMB #0970-0506), the team achieved a 90 percent response rate to a web-based survey administered to TANF staff,

which used the same mode and targeted a similar population as planned for the NextGen Project. These surveys are designed to be easy to complete, use straightforward language, and allow respondents to break off and complete later if they get interrupted. To maximize response rates and data reliability, periodic email reminders will be sent to respondents, beginning two weeks after the field period begins. If staff and leaders do not complete their surveys within one week of the targeted time frame, the designated liaison will follow up with the site point of contact to remind staff and leaders that survey responses are due.

- **Semi-structured discussions with program staff and employers.** The project team will target completing 200 semi-structured discussions with staff (an average of 20 at each program) and 50 semi-structured discussions with employers (an average of 5 at each program). The project team expects program leaders to support this data collection, greatly reducing the risk of nonresponse. To maximize participation, well before the site visits during which the semi-structured discussions will take place, the project team will begin scheduling with program staff and employers to ensure the timing is convenient.
- **In-depth participant interviews.** The respondents to these interviews will be a convenience sample. The project team will target completing interviews with approximately 20 treatment group members from each program. Because they want a convenience sample, the team will reach out to randomly selected treatment group members until they have agreement to participate in the interviews by about 25 people. Based on experiences recruiting for the Evaluation of Employment Coaching for TANF and Related Populations, this will be sufficient for enough people to show up for the interviews to interview 20 people. The project team will be flexible in scheduling interview times and locations to accommodate study participants' schedules and needs.

Cost studies

Based on similar research projects, such as the Evaluation of SNAP Employment and Training Pilots (OMB #0584-0604) and the WIA Adult and Dislocated Worker Programs Gold Standard Evaluation (OMB #1205-0504), the project team expects all programs to provide cost data. To maximize responses, the project team will be flexible with the data collection approach, asking programs to submit their existing accounting records and tailoring the sections of the cost workbook to fill in any gaps after reviewing those records. The project team will also provide technical assistance to the programs as they complete the workbook.

Nonresponse

Impact Studies

During survey fielding for the first and second follow-up surveys, the team will actively monitor response rates, with an eye to any treatment–control differences. If such differences are observed, the project team will intensify the locating efforts for the group with the lower response rate to minimize differential nonresponse during active data collection.

Following the close of data collection, the project team will analyze nonresponse on the follow-up surveys to assess whether the survey respondents are representative of the full study sample. Using the data on participants' characteristics collected at baseline, the project team will conduct statistical tests (chi-square and t-tests) to gauge whether the treatment group members who

participated in data collection are representative of all the treatment group members, whether the control group members who participated in data collection are representative of all the control group members, and whether there are systematic differences in the treatment and control group members who responded to the survey.

The project team will use two approaches to correct for potential nonresponse bias in the estimation of program impacts. First, the regression models described in Section B7 will adjust for observed differences between the characteristics of treatment and control group respondents. Second, because this regression procedure will not correct for differences between respondents and nonrespondents in each research group, the project team will construct sample weights so that the weighted baseline characteristics of respondents in the treatment and control group in each program are similar to those of the full sample (respondents and nonrespondents). These weights will be constructed using data from the baseline surveys.

To reduce any bias resulting from item nonresponse, the project team will impute values for missing data. Imputation is particularly useful in cases in which data might be systematically missing related to an observable characteristic. For example, if a study participant was not employed, the team knows that his or her wage and salary earnings will be zero. However, many more data items are required to construct a measure of earnings for employed individuals and, thus, it is more likely that employed individuals will have missing earnings. This suggests that, without imputation, estimates of earnings might be biased downward. The imputation approaches used will include logical imputation, predictive mean matching, and hot-deck imputation. The approach used will be determined by the type of data that are missing.

Descriptive and Cost Studies

The project team will not statistically adjust the data collected for the descriptive and cost studies for nonresponse, as the goal is to produce comprehensive, accurate information about the program and its costs rather than calculate statistics about a population.

B6. Production of Estimates and Projections

The estimates from this project will be released publicly following ACF review.

Impact studies. The impact studies will estimate the effectiveness of each program in the study in improving outcomes of study participants. Any observed differences in outcomes between the treatment and control group members can be attributed to the effectiveness of the program; in statistical terms, the differences are internally valid estimates of the mean impacts of the program, as delivered, on the corresponding outcomes for similar populations in the same environment.

The project team will use the constructed sample weights described in Section B5 in the impact analysis so that the weighted baseline characteristics of respondents in the treatment and control group in each program are similar to those of the full sample (respondents and nonrespondents). The project team will also address missing responses as described in Section B5.

The baseline data will be used to describe the study participants in each program. The project team will use chi-squared tests of differences in means over all characteristics to assess whether random assignment successfully generated treatment and control groups with similar baseline characteristics, and that the treatment and control group respondents to the follow-up surveys are similar. The project team will also report t-tests of differences for individual characteristics.

Impacts will be estimated for each program. The project team will use regression estimators to control for residual differences between the treatment and control groups and to construct more efficient estimators than the simple difference-in-means estimators (as described below).

Descriptive studies. The data collection for the descriptive studies will not result in statistical estimation or projections. The evaluation of each program will describe how the program was designed and implemented, the program's environment, challenges the programs faced implementing the program and how they were addressed, the program's potential sustainability, who participates, duration of participation, and the program services participants received. This information will help interpret the impact findings and enable other programs to replicate the program.

Cost studies. The data collection for the cost studies will not result in statistical estimation or projections. The project team will use information reported by the programs in the Excel-based workbooks to estimate an average cost of the program per participant-month. In addition, the team will use data collected from the surveys and administrative records to estimate the average benefit of the program per participant and compare the benefits and costs.

B7. Data Handling and Analysis

Data Handling

Survey data. The web survey and the telephone interview software will use real-time logic rules, enforce skip patterns, and provide soft and hard checks. Soft and hard checks will be displayed for interviewers or respondents if the provided information conflicts with earlier responses or is out of range for expected values. Hard checks require resolution before continuing; soft checks can be suppressed. All CATI interviewers are subject to real-time monitoring to ensure they are correctly interpreting and entering respondent responses. Following data collection, the project team will conduct comprehensive data reviews and quality assurance reviews to ensure skip patterns are enforced and data are complete and within expected ranges.

During data processing and coding, the project team will conduct quality assurance reviews to ensure consistency and minimize any data processing errors. Specifically, coders will participate in a comprehensive training session, and the project team will monitor their work, perform quality control checks, and conduct quality assurance reviews of all weighting and imputation procedures. Any outliers, skip logic errors, or other recodes of survey data will be recorded in both internal programs and data editing spreadsheets.

RAPTER®. As with the survey software, RAPTER® uses real-time logic rules and validity checks to prevent entry errors. The project team will extensively test all functionality. The project team will train all program staff in the use of the system, provide them with a written procedures manual, and routinely examine the data staff enter into RAPTER® to ensure quality.

Interview data from discussion guides for the staff, employer, and in-depth participant interviews. The project team staff will be trained on instruments for the descriptive studies and how to appropriately code that information after collecting it. During data coding, quality assurance reviews will be conducted to ensure consistency. The descriptive study task leader or other senior member of the project team will perform quality control checks of the coded information.

Data Analysis

Impact studies. The impact analysis will consist of comparisons of means and distributions using the randomly assigned treatment and comparison group. This will include unadjusted and regression-adjusted means for outcomes including employment, earnings, and benefit receipt. Differences in means or proportions of follow-up outcomes between the treatment and control group will provide unbiased estimates of the impacts of the program. Estimates that are more precise will be obtained using regression models to control for random differences in the baseline characteristics of treatment and control group members. In their simplest forms, these models can be expressed by the following equation: $Y_i = \alpha + \beta X_i + \delta T_i + \varepsilon_i$, where Y_i is an outcome (such as earnings) for person i; α is a constant; X_i is a (column) vector of baseline characteristics (such as sex, 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 treatment; δ represents the impact of the program; and ε_i is an error term. These models will be estimated separately for each program.

If the sample is large enough, the project team will conduct a subgroup analysis to examine who benefits most from the program. Subgroup effects will be estimated using the following equation:
$Y_i = \alpha + \beta X_i + \delta_1 T_i + \delta_2 G_i + \delta_3 T_i G_i + \varepsilon_i$, where G_i is an indicator for whether person i is part of a subgroup; δ_2 represents the relationship between subgroup status and the outcome; and δ_3 represents the additional effect of treatment for those in the subgroup. The project team will consider subgroups that are appropriate for the program's target population, such as those defined by disability status, work readiness, employment challenges, or history of TANF receipt.

Descriptive studies. To analyze the large amount of interview data collected from multiple sources efficiently and accurately, the project team will develop a coding scheme that maps to the implementation framework, research questions, and programs' theories of change. After coding the data, the team will look for common themes across data elements and respondents and examine the extent to which the programs adhered to the fidelity measures defined during the program selection phase. The analysis will include an assessment of conditions needed to replicate and sustain the program. To analyze the information from the staff and leadership surveys, the team will compute descriptive statistics, such as means and frequencies.

Cost studies. The project team will use information from the cost studies to compute an average cost of the program per participant-month. From this, and using information on the average number of months the program participants in the study were engaged in the program, the average cost of the program per participant will be estimated. The team will use data collected from the surveys and administrative records to estimate the average benefit of the program per

participant and compare the benefits and costs. Doing so involves considering the possible benefits and costs of the program from the participants', government, and rest of society's perspectives.

Data Use

The project team will publish findings from the project throughout the study in technical reports and briefs. The project team anticipates that reporting on the descriptive and cost studies will occur in 2021 and 2022; reporting on the intermediate impact findings will take place in 2022 and 2023; and reporting on the final impact findings will occur in 2023 and 2024. In addition to presenting findings, reports and briefs will document the methodologies used to collect, process, and analyze the project's data across the impact, descriptive, and cost studies; this will assist readers in assessing study quality and interpreting the findings. Study limitations and information about the generalizability of the results will be included when presenting findings.

In addition, the project will prepare final data files and documentation to be available publicly so other researchers are able to duplicate all analyses. The provided documentation will improve the understanding of how to properly interpret, analyze, and evaluate the information resulting from the data collection. The project team anticipates that data archives (restricted or public use) would become available in 2024 and hosted on a data archive platform such as the Inter-university Consortium for Political and Social Research (ICPSR).

B8. Contact Person(s)

Contact information for people who can answer questions about the statistical aspects of the survey:

- Hilary Bruck: hilary.bruck@acf.hhs.gov
- Sheena McConnell: smcconnell@mathematica-mpr.com

Mathematica developed the plans for this data collection. Leaders of the project team from OPRE, Mathematica, and Tree House Economics who designed and/or will collect and analyze the data are as follows:

- Hilary Bruck, senior social science research analyst, ACF
- Gabrielle Newell, social science research analyst, ACF
- Marie Lawrence, social science research analyst, ACF
- Sheena McConnell, senior vice president, Mathematica
- Michelle Derr, senior researcher, Mathematica
- David Stapleton, partner, Tree House Economics
- Annalisa Mastri, senior researcher, Mathematica
- Jody Schimmel-Hyde, senior researcher, Mathematica
- Kristen Joyce, researcher, Mathematica
- Ryan Callahan, survey researcher, Mathematica

Attachments

Instruments

Instrument 1. Baseline survey
Instrument 2. Identifying and contact information

Appendices
Appendix A. Informed consent form
Appendix B. Question-by-question justification for baseline survey
Appendix C. Question-by-question justification for identifying and contact information
Appendix D. Question-by-question justification for follow-up surveys
Appendix E. Reporting burden and cost for Phase 2 data collection instruments
Appendix F. Instrument 3 (draft): First follow-up survey
Appendix G. Follow-up survey reminders and notifications
Appendix H. Instrument 4 (draft): Second follow-up survey
Appendix I. Instrument 5 (draft): Service receipt tracking
Appendix J. Instrument 6 (draft): Staff characteristics survey
Appendix K. Instrument 7 (draft): Program leadership survey
Appendix L. Instrument 8 (draft): Semi-structured program discussion guide
Appendix M. Instrument 9 (draft): Semi-structured employer discussion guide
Appendix N. Instrument 10 (draft): In-depth participant interview guide
Appendix O. Instrument 11 (draft): Cost workbook
Appendix P. Federal Register Notice

Supporting Statement B References

What Works Clearinghouse. "Standards Handbook, Version 4." 2017. Available at https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf.