



MATHEMATICA
Policy Research, Inc.

**Green Jobs and
Health Care
Impact Evaluation**

**Evaluation
Design Report**

Final

December 2011

Prepared for:

Savi Swick

U.S. Department of Labor, ETA
200 Constitution Avenue,
NW N-5641
Washington, DC 20210

Submitted by:

Abt Associates Inc.

4550 Montgomery Avenue
Suite 800 North
Bethesda, MD 20814

In Partnership with:

Mathematica Policy Research,
Inc.

P.O. Box 2393
Princeton, NJ 08543-2393e

Table of Contents

Section 1 – Introduction 1

- Background on GJ-HC Impact Evaluation 1
- Research Questions that the GJ-HC Impact Evaluation Addresses..... 2
- Conceptual Framework 3
- Overview of the Report 5

Section 2 – Site Selection 6

- 2.1 Site Selection Framework and Process 6
 - 2.1.1 Programmatic Criteria 6
 - 2.1.2 Technical Criteria..... 6
 - 2.1.3 Research Capacity Criteria..... 7
 - 2.1.4 Selection Process..... 7
- 2.2 Selected Grantees 8

Section 3 – Developing and Implementing Random Assignment for the Evaluation 12

- 3.1 The Random Assignment Process..... 12
 - 3.1.1 Developing Site-Specific Processes 13
 - 3.1.2 Site-Specific Manuals 14
 - 3.1.3 Site Training..... 15
- 3.2 Baseline Data Collection..... 15
 - 3.2.1 Participant Consent Form (PCF)..... 15
 - 3.2.2 Baseline Information Form (BIF) 16
 - 3.2.3 Data Quality Control 16
- 3.3 Randomization Procedures..... 18
 - 3.3.1 Goals and Properties of Random Assignment..... 18
 - 3.3.2 Stratification, Pre-Designation, and Block Size 18
 - 3.3.3 Randomization through a Computer Platform 20

Section 4 – Data Collection..... 22

- 4.1 Survey Data..... 22
 - 4.1.1 Baseline 22
 - 4.1.2 Follow-up Surveys 22
 - 4.1.3 Survey Data Collection Procedures..... 26
- 4.2 Administrative Data on Employment and Earnings..... 30
 - 4.2.1 State UI Wage Records 30
 - 4.2.2 National Directory of New Hires 31
 - 4.2.3 Collecting UI Data 31

4.3	Data Collection for the Process Study	32
4.3.1	Interviews with Program Staff and Partners	32
4.3.2	Program Documents.....	33
4.3.3	Program Enrollment, Attendance, and Completion Data.....	34
4.3.4	Participant Focus Groups	34
Section 5 – Process Analysis.....		35
5.1	Research Areas.....	35
5.2	Key Program Dimensions	35
5.2.1	Program Design and Operations	36
5.2.2	Local Context.....	37
5.2.3	Service Receipt and Utilization.....	37
5.2.4	Participant Views of Services and Barriers.....	37
5.2.5	Implementation Accomplishments and Challenges	37
5.3	Use of Data Sources	38
5.4	Analysis Methods.....	39
Section 6 – Impact Analysis		40
6.1	Impact Research Questions	40
6.1.1	Primary Confirmatory Outcome.....	40
6.1.2	Data Sources for the Confirmatory Outcome Measure	43
6.1.3	Exploratory Analyses	45
6.1.4	Principal Focus on the Effects of Access to Training	47
6.2	Minimum Detectable Impacts (MDIs).....	47
6.3	Estimation Methods	51
Section 7 – Project Activities and Schedule		53
7.1	Task 1: Selection of Sites.....	55
7.2	Task 2: Evaluation Design	55
7.3	Task 3: Process Study	55
7.4	Task 4: Implementation and Monitoring of Random Assignment.....	56
7.5	Task 5: Follow-Up Surveys	56
7.6	Task 6: Preparation of Reports.....	57
7.7	Task 7: Peer Review Panel.....	59
7.8	Task 8: Oral Briefings.....	59
7.9	Task 9: Administrative Data Collection and Protection of Personally Identifiable Information	59
7.10	Monthly Progress Reports and Management	60
Works Cited.....		61

Section 1 – Introduction

This report presents the design for evaluating the U.S. Department of Labor’s (DOL) Green Jobs and Health Care (GJ-HC) Impact Evaluation, including the impact analysis and the process study. This report (1) reviews the overall research objectives and key questions that the GJ-HC process and impact studies address; (2) lays out the plan and main tasks involved in data collection and analysis; and (3) provides a proposed schedule for completion of major project tasks and reports. The infusion of new funding to support training for green jobs and health care occupations provides an opportunity to test the extent to which of these new programs improve worker outcomes by imparting skills and training valued in the labor market.

Background on GJ-HC Impact Evaluation

The Green Jobs and Health Care (GJ-HC) Impact Evaluation addresses one of the central challenges facing policymakers and administrators in this period of economic decline: How effective are specific strategies designed to help individuals obtain and succeed in training in select industries at promoting sustained employment and advancement in the labor market? The Employment and Training Administration (ETA) at the U.S Department of Labor (DOL) is addressing this vital issue by providing resources to develop training programs in high-growth fields, particularly the energy efficiency and renewable energy and health care fields and conducting a rigorous, multi-site evaluation to build a firm base of knowledge to inform future policy.

As part of a comprehensive economic stimulus package funded under the 2009 American Recovery and Reinvestment Act (ARRA), DOL unveiled a series of grant initiatives to promote training and employment in select high growth sectors of the economy, explicitly “energy efficiency and renewable energy” and the “health care sector” per the Act’s final language. Several grant programs emerged from this legislation, including ones for capacity building among training providers, labor information management, and for the development of statewide labor partnerships. Two programs providing training primarily in the energy efficiency and renewable energy and health care fields are the focus of this evaluation: the Pathways Out of Poverty (Pathways) and Health Care and Other High Growth Industries (Health Care) initiatives. While each grant initiative has its own unique set of objectives, their common aim is to help prepare workers with the skills needed to successfully pursue jobs and career opportunities in select growth sectors. Key features of each grant initiative are:

- ***Pathways Grantees.*** DOL funded 38 Pathways Out of Poverty grantees to provide training and placement in “green” occupations with an emphasis on the energy efficiency and renewable energy sectors. The funded programs target economically disadvantaged populations with a focus on high-poverty regions. DOL awarded these two-year grants in January 2010.
- ***Health Care Grantees.*** The Health Care and Other High Growth Industries initiative funds grantees in any geographic area regardless of regional poverty level. The grants focus on nursing, allied health, long-term care, and health information technology. In addition, but to a lesser extent, the grants fund training in other “high growth, high demand, and economically vital sectors of the American economy” such as information technology, advanced manufacturing, and biotechnology. DOL awarded 55 three-year grants under this funding stream in February 2010.

In October 2010, DOL/ETA selected Abt Associates, Inc. and Mathematica Policy Research, Inc. to conduct the Green Jobs and Health Care Impact Evaluation. As specified by ETA, the evaluation called for an experimental research design using random assignment in selected sites to evaluate the two grant

programs. The evaluation has two primary goals: (1) to estimate the selected programs' impacts on employment, earnings, and career advancement, and (2) to identify promising strategies for replication. In particular, the evaluation will identify the extent to which selected programs have positive impacts but will not seek to determine whether the overall ARRA grant funding streams are effective public policy interventions.

This Evaluation Design Report describes our plans for the evaluation, specifically the research questions to be addressed and the data and analytic methods to be used to answer those questions. The site selection process is complete, with a total of four grantees selected across the Pathways and Health Care programs. The selected grantees began to implement random assignment in August 2011. The balance of this section presents the goals of the evaluation and describes the structure of the remainder of the report.

Research Questions that the GJ-HC Impact Evaluation Addresses

The evaluation examines both the implementation and impact of training services funded by the two ARRA grant programs in four selected sites. The research questions that the evaluation addresses are:

- What is the impact of each selected grantee's program on participants' receipt of education and training services, in terms of both the number who receive these services and the total hours of training received?
- What is the impact of each program on the completion of educational programs and the receipt of certificates and credentials from the training?
- What is the impact of each program on the employment levels and earnings? To what extent does each program result in earnings progression?
- To what extent does each program result in any employment (regardless of sector)? To what extent does each program result in employment in the specified sector in which the training was focused?
- What features of the programs are suggestive of impacts, particularly in terms of target group, curricula and course design, employer connections, and additional supports?
- What are the lessons for future programs and practices?

Because the evaluation will use a rigorous experimental research design, the evaluation examines the extent to which *adding* the GJ-HC training services in each site to *the configuration of training services already available in the community* improves participant outcomes. This focus on the incremental contribution of each grant program to its community, through a comparison of the program's services to other services in the community, is a common design for studies of training programs, and the one that tells policy makers what they need to know—whether funding the type of interventions studied makes a difference in participants' educational and employment outcomes, compared to already-existing employment and training services. Because study subjects seek out job training on their own, evaluation findings will not apply to all individuals eligible for program services in a given site. The participants actually studied are more likely to be more motivated to succeed in training and the labor market than the broader eligible population and thus could experience impacts that differ from what would be seen if the entire eligible population were induced to participate. However, findings will accurately characterize program impacts on individuals who voluntarily seek out training.

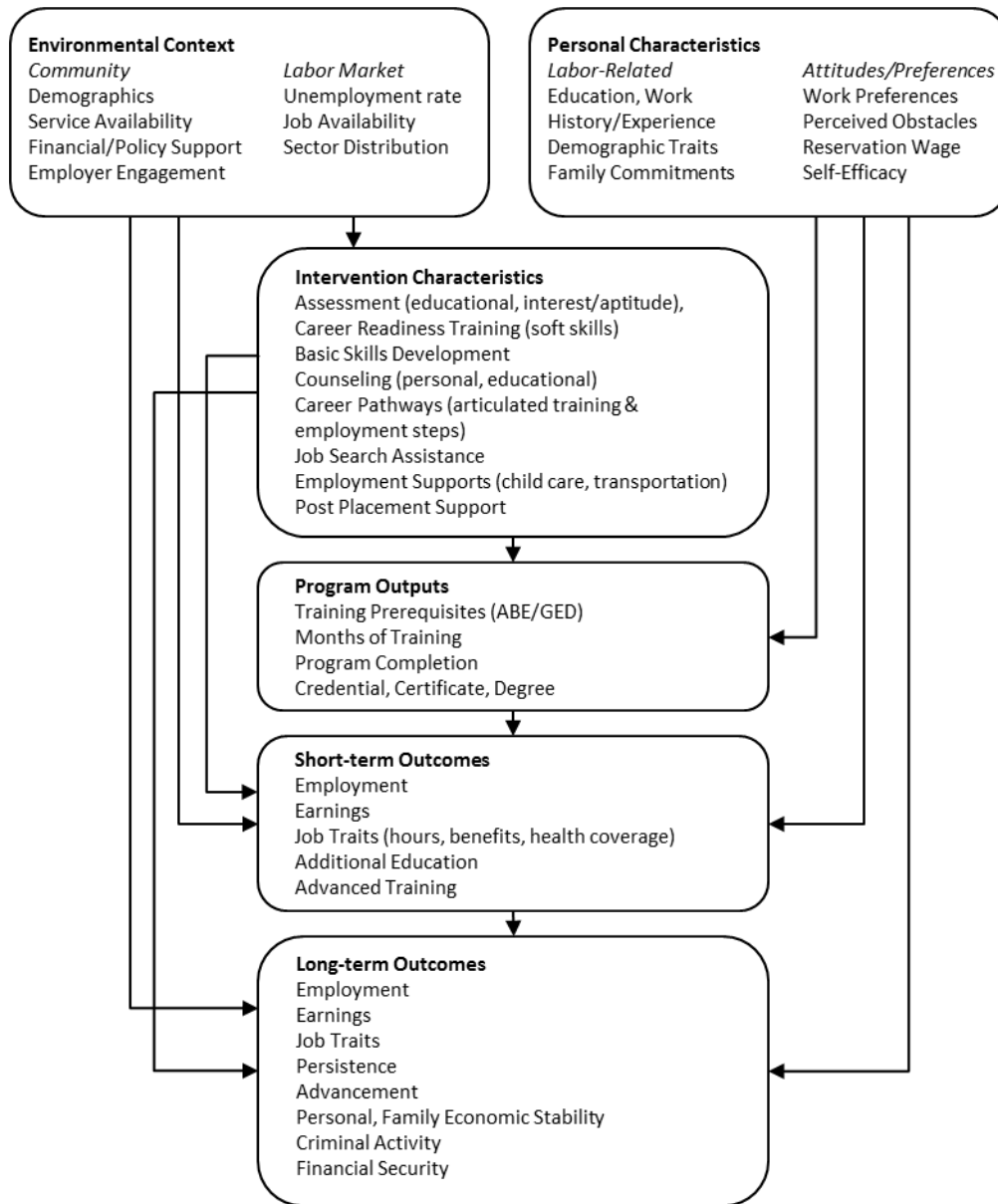
Conceptual Framework

Our evaluation plan is motivated by the conceptual framework depicted graphically in Figure 1.1. Specifically, the framework expresses our view of how elements of the intervention as well as outside factors influence short- and long-term outcomes. The “intervention characteristics” are what DOL is funding under the GJ-HC grants. The programs themselves generate outputs as the box below notes. These program outputs then lead to the outcomes of employment, earnings, job quality, potential additional education and advanced training. These short-term outcomes lead to the longer term outcomes of potentially better employment, earnings, job quality, job persistence, career advancement, and personal or family economic stability. The model recognizes that it is not only the program’s characteristics that generate these outputs and outcomes but also that environmental context and personal characteristics are influential factors. The arrows between specific boxes in the model represent the expected influences among the factors. Although we present a consolidated conceptual model here, we expect the process evaluation to document each site’s particular conceptual framework, also described as a logic model or theory of change.

To elaborate on one possible path of influence, for example, consider the box that contains personal characteristics. We expect that personal characteristics—both labor-related and attitudinal—will influence the ways in which participants interact with the intervention itself. Further, their personal characteristics will also influence the expected outputs and outcomes that arise from program participation. People with stronger work and educational backgrounds and more flexibility in the types of work they are willing to do will have favorable post-program employment outcomes.

The conceptual framework design is general enough to capture variation along several dimensions. A major dimension of interest is what distinguishes green jobs from health care or high-tech industry jobs; and elements of the conceptual framework capture variation that might exist in that regard. For example, among environmental factors that might matter, we include sector distribution, which might be measured practically as the percent of the local labor market that manufacturing jobs comprises. Similarly, what is included in the central Intervention Characteristics box is site-specific, as relevant to the evaluation, but can all be captured in one overarching framework. This conceptual framework is foundational for both the process and impact portions of the evaluation. For instance, the process analysis will create site-specific versions of this conceptual model, providing rich context for use in interpreting results from the impact analysis; and for the impact analysis, the model suggests some key subgroup analyses to explore.

Figure 1-1. Conceptual Framework for GJ-HC Impact Evaluation



Overview of the Report

The remainder of this Evaluation Design Report is organized as follows. Section 2 discusses our site selection framework, process, and results. The section concludes with a description of the selected grantees. Section 3 discusses the plan for random assignment. In particular, the section considers the relation between sites' participant intake process and how the evaluation will implement random assignment. Section 4 discusses the evaluation's sources of data and data collection procedures. Section 5 describes the research questions, data sources, and analytic methods for the process analysis. Section 6 does the same for the impact analysis. Finally, Section 7 describes the content of the evaluation's reports and project timeline.

Section 2 – Site Selection

Selecting the grantees to be included in the evaluation is a critical element of the evaluation’s research design. This step of the evaluation design is complete and described below. In coordination with DOL, the evaluation team first developed a site selection framework that specified the criteria for selecting grantees. Using this framework, we then selected sites that have a strong intervention, allowing the evaluation to address areas of interest to DOL; the size and scale to meet the statistical requirements of the evaluation, and the capacity to implement a random assignment evaluation. This section discusses the site selection framework and process and the results that identified the four sites to be part of the evaluation.

2.1 Site Selection Framework and Process

Our site selection process considered three criteria: (1) programmatic criteria that account for the nature of the grantees’ interventions; (2) technical criteria that emphasize the statistical requirements of the evaluation design; and (3) research capacity criteria that address the sites’ ability to implement a random assignment study. We discuss each of these in turn.

2.1.1 Programmatic Criteria

In initial meetings about the evaluation, DOL and the evaluation team agreed to focus on evaluating programs with innovations that make it more likely that low-skilled individuals would complete training and obtain credentials. Specifically, we agreed that the evaluation would focus on innovative training programs that move beyond “business as usual” to provide comprehensive training services including accommodations and supports to more effectively serve low-skilled individuals. Programs of interest had elements of a “career pathway” approach that fosters job advancement through articulated training and employment steps in occupations in demand in local communities. This approach to job training seeks to provide a set of comprehensive services with a number of programmatic elements that improve training outcomes generally, and particularly those for low-skilled individuals. These elements can include instructional accommodations, particularly the inclusion of basic skills instruction in addition to training; supportive services such as career and personal counseling, soft skills, and financial assistance; and building connections to employers and jobs.

Including as a programmatic criterion the comprehensiveness of the grantee’s intervention has several strengths. While strong interest in career pathway approaches exists nationally, limited rigorous research has examined the effects of these types of programs; therefore, this evaluation can build needed knowledge about these types of training strategies. Further, a strong distinction between the services that the treatment and control groups receive is likely. Finally, grantees integrating these design features can be included regardless of industry, allowing green jobs grantees and some health care grantees to be included in one unified framework.

2.1.2 Technical Criteria

We considered technical criteria, associated with ensuring that the size and composition of the sample were adequate to detect differences in outcomes between the treatment and control groups, if they exist. Three issues were important in this area: program size, enrollment flow and timing, and the type and size of services available to the control group.

- **Program size:** The evaluation seeks to generate site-specific impact estimates. To do so, the evaluation team estimated that sample sizes of approximately 600 people per group are desirable,

with minimum sample sizes of around 450 needed to detect site-specific impacts of a policy-relevant size with adequate certainty.

- ***Enrollment flow and timing:*** Related to the overall program size criterion are considerations of flow and timing of the enrollment cycles. Of particular importance is the extent to which the flow of applicants is adequate to achieve sample size targets in the original 12-month timeframe designated for random assignment. (As discussed below, a random assignment period of up to 18 months is likely to be needed in some sites to meet sample size goals).
- ***Creating a control group.*** In regard to creating the control group, two separate issues were important. The first issue concerned the grantee’s ability to “over-recruit” and both fill program slots and create a control group. The second issue concerned a preference for sites where nothing similar to the grant-funded training under study exists already in the community that the control group could access. We would not expect to learn much about the effectiveness of these services if treatment and control group members receive them at similar rates, albeit from different sources. Sites in which the treatment-control contrast is sharper provide better opportunities for learning about the effectiveness of program services.

2.1.3 Research Capacity Criteria

A successful experimental evaluation requires forming a working partnership between the evaluation team and participating grantees. At a minimum, sites selected for the study must be fundamentally supportive of the basic research mission and methodology of the study including:

- ***Willingness to participate:*** Although most sites are fully supportive of DOL’s research agenda, concerns exist about random assignment in a program setting. While the grant award stipulates that grantees must participate in evaluation activities, it did not specify that these activities would involve random assignment. Some believe it is unethical to “withhold services” from individuals in need, while others think it is politically unwise to do so. Ultimately, grantees selected to participate in the study had to be willing and able to overcome these concerns so that they are not an impediment to a multi-year study.
- ***Willingness to accommodate the study:*** In addition to a broad willingness to participate, grantees must be able to accommodate the needs of such an evaluation. While the research team always seeks to limit site burden, it is necessary to modify operations to some extent in order to consistently implement the study. Such modifications entail the collection of new data or the use of a slightly modified intake process to allow for random assignment to be made at the proper point.

The evaluation team used these programmatic, technical, and research capacity criteria as the basis for the site selection process described below.

2.1.4 Selection Process

Below we describe the key activities that comprised the site selection review process.

Review of all grant applications for all awarded grants: The research team reviewed all 93 grant applications that DOL funded under the Pathways Out of Poverty and the Health Care and Other High Growth Industries programs. We focused on identifying programs that both provide comprehensive training featuring elements of career pathways and operate at a scale sufficient for the study.

Telephone calls to a sub-sample of sites pre-selected by DOL. Based on this initial review as well as some recommendations from DOL, the site selection team conducted a series of in-depth phone interviews with the program directors and other key staff from 30 grantees that had potential for meeting the criterion for inclusion in the evaluation. Each interview was approximately 60-75 minutes in length and gathered detail on the nature of the training programs, current and planned program enrollment, program operations, and implementation progress. During these calls, the team also gauged the grantees' preliminary receptivity to participation in the study and assessed what, if any, barriers would exist. From these 30 total interviews, the team identified ten sites that best met the selection criteria.

Site visits to grantees: To gather more detailed information and further assess sites' appropriateness for the project, one-day site visits took place with nine of the grantees, and a series of conference calls took place with the tenth (due to scheduling/logistical issues). These site visits examined the nature and intensity of the service delivery model; the grantees' capacity to meet enrollment targets required by the evaluation; the suitability of the enrollment process for random assignment; barriers to the implementation of a random assignment design; the ability to over-recruit for the control group; and the availability of community training and support services available. They also included discussions with grantee stakeholders about possible participation in the study.

Final recommendations and selection: Based on the site visits, the evaluation team recommended four sites for inclusion in the evaluation. We provided an overview of the findings and recommendations from the site visits to DOL, whose staff concurred with the recommendations.

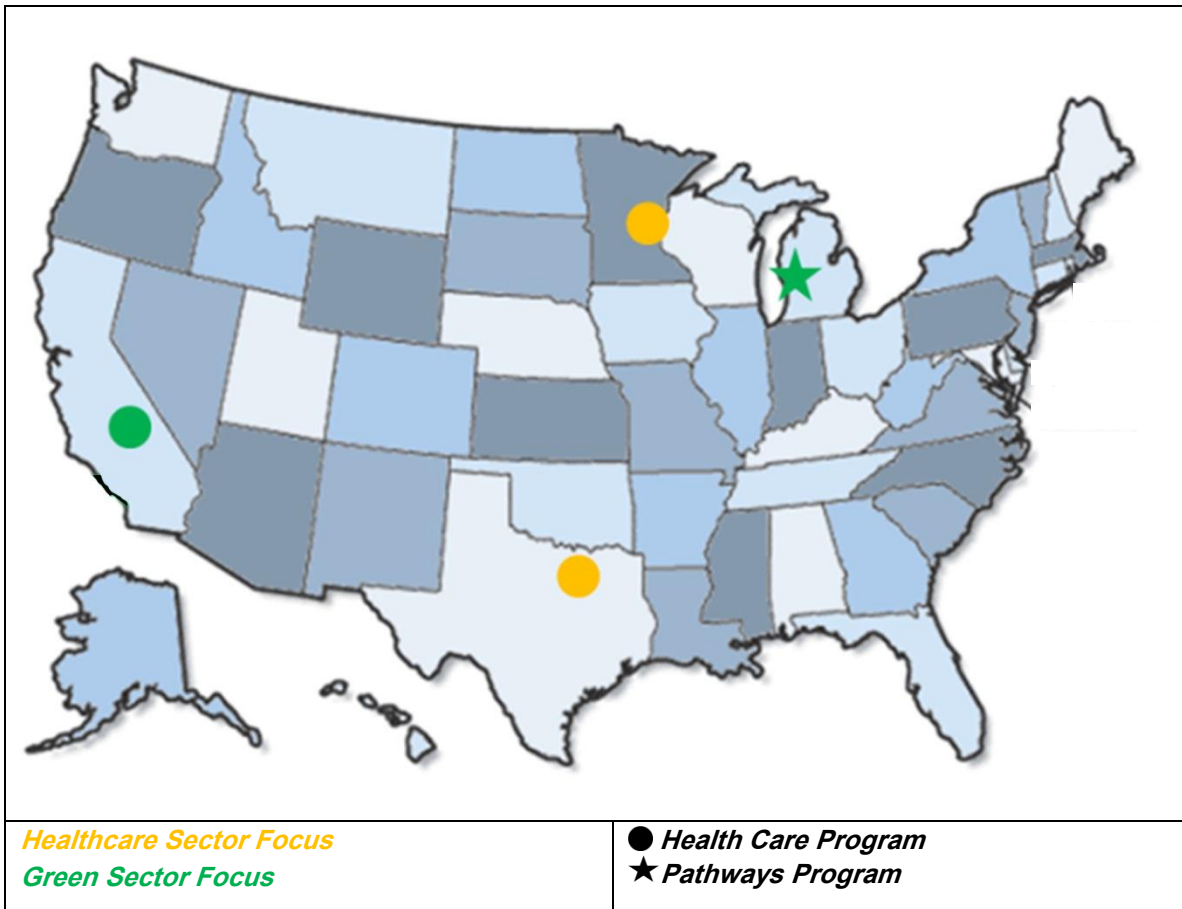
2.2 Selected Grantees

One Pathways Out of Poverty and three Health Care and Other High Growth Industries grantees satisfy the primary criteria for the study. They are:

- Grand Rapids Community College (Grand Rapids, MI)
- Kern Community College District (Bakersfield, CA)
- American Indian Opportunities Industrialization Center (Hennepin County (Minneapolis), MN)
- North Central Texas College (serving Gainesville, Corinth, Flower Mound, Bowie and Graham, TX)

Figure 2-1 illustrates the grantees' locations and shows that two of the four included grantees focus on "green" sector training and the other two on healthcare sector training. Figure 2-2 briefly summarizes each site's intervention and expected sample size.

Figure 2-1. Map of Grantees



The training programs that this evaluation will study are those programs that are feasible to evaluate with adequate sample sizes using a random assignment design. They do not represent “best practices” and are not selected to be representative of all ARRA-funded programs under these two grant streams. To make this clear and explicit to future readers of evaluation reports, we will discuss in those reports how each of these four programs compares to the full set of 93 grantees funded by the two ARRA solicitations. This will provide important context for interpreting the study’s results, allowing us to emphasize that the findings are not representative of the two grant programs as a whole but that they provide information on the effects of four distinct interventions funded by those programs.

Figure 2-2. Summary of GJ-HC Grantee Programs Recommended for Impact Evaluation

Program/ Expected Sample Size	Sector/ Certifications	Target Group	Program Description
<p>Grand Rapids Community College (GRCC)</p> <p>Grand Rapids, MI</p> <p><i>(Pathways Out of Poverty Grantee)</i> 600 treatment group members</p> <p>300 control group members</p>	<p>Green sector</p> <p>(National Career Readiness Certificate; Michigan Employability Certificate, Green Advantage (for building trades); and various industry certificates)</p>	<p>Although the program does not target any specific groups, there are many displaced workers in Central Western Michigan. Additionally, since the program does not prohibit ex-offenders from applying, many participants have a criminal record.</p>	<p>Grand Rapids Community College (GRCC) integrates existing and new education/job training, placement, retention and support service programs to assist the target populations with attaining and retaining employment in high growth green industry occupations in the region. After assessment, students develop an educational plan that outlines their career path and delineates steps to get there. An initial six to eight week training results in a college-readiness certificate and includes an inquiry-based developmental science course; basic math, reading, and locating information skills; and employability skills. Once students complete initial training, they have the option to begin work immediately or go on to specific training in a particular field (e.g., deconstruction, green building, solar, or water science). Students who enter with more advanced skills may be able to bypass the six to eight week introductory training. Occupational training can last between 80 hours and two years. Students are paired with a career counselor who meets with them throughout the program. Additional supportive services (e.g., transportation subsidies; funds for books, supplies, and tools) can be made available by the program.</p>
<p>Kern Community College District (KCCD)</p> <p>Bakersfield, CA</p> <p><i>(Health Care and Other High Growth Industries Grantee)</i></p> <p>425 treatment group members</p> <p>425 control group members</p>	<p>Green sector</p> <p>(Introductory track: KCC completion certificate; Solar technician track: North American Board of Certified Energy Professionals (NABCEP) certification; and, Wind technician track: KCC completion certificate (an industry standard training certification does not exist))</p>	<p>In general the target population is dislocated workers and unemployed individuals. Participants must have their high school diploma or GED, a driver's license, a score of 4 on Work Keys for reading, math and locating information, a clean drug test, and WIA eligibility for unemployed and underemployed individuals.</p>	<p>Kern Community College District (KCCD) trains unemployed, dislocated, and incumbent workers for technician and construction employment in the renewable energy industry. Participants complete training in three renewable energy industry occupations: utility worker (an entry-level track), wind technician, and solar technician. All participants begin with the Power Tech course where they receive foundation training as utility workers. This is a pre-requisite for the other courses. This course lasts five weeks. Upon completion, participants can leave the program and search for jobs, or stay on and continue into the solar power or wind power tracks. In the solar power track, students learn to install and repair solar power systems. The track consists of one course that lasts for eight weeks. In the wind power track, students learn how to install, repair and clean wind turbines, with a special emphasis on on-the-job safety. The track consists of one course lasting eight weeks. All three programs meet daily for two three-hour modules. Programs include field trips to local companies and projects that allow participants to receive first-hand experience and work with industry equipment during their training. Courses are taught by industry professionals. Support services include necessary equipment, uniforms, childcare, eye exams, and transportation.</p>

Figure 2-2. Summary of GJ-HC Grantee Programs Recommended for Impact Evaluation (continued)

Program/ Expected Sample Size	Sector/ Certifications	Target Group	Program Description
<p>American Indian Opportunities Industrialization Center (AIOIC)</p> <p>Minneapolis, MN</p> <p><i>(Health Care and Other High Growth Industries Grantee)</i></p> <p>600 treatment group members</p> <p>600 control group members</p>	<p>Healthcare sector</p> <p>(Certified Nursing Assistant (CNA); Medication Aide, Home Health Aide; Acute Care Specialist; and EMT)</p>	<p>Participants must be at least 18 years old, have no criminal record, and pass at least the 7th grade math and reading level (passing at the 6th grade level is acceptable for some programs). It is not necessary for participants to have their GED.</p>	<p>American Indian Opportunities Industrialization Center (AIOIC) offers courses that are part of a career pathway in the nursing sector, providing training for several lower-level certificates in the health sector. Students have the option of taking short-term classes individually before seeking employment, or enrolling in a longer-term (six to nine month) training program. The six-month program combines the short-term classes into a streamlined progression leading to Trained Medication Aide certification. The nine-month program leads to Acute Care Specialist certification and includes the same coursework as the six-month program with additional specialized courses. These courses form the basis of the prerequisites for enrolling in a Licensed Practical Nurse (LPN) or Registered Nurse (RN) program. Under the grant, AIOIC offers supportive services including: a self-empowerment program to address personal circumstances like parenting, personal finance, housing, health, domestic violence, legal issues, and cultural studies; academic advising and financial counseling; computer labs; additional financial assistance for training and certification-related expenses, and comprehensive placement and post-employment services.</p>
<p>North Central Texas College (NCTC)</p> <p>Gainesville, TX</p> <p><i>(Health Care and Other High Growth Industries Grantee)</i></p> <p>589 treatment group members</p> <p>485 control group members</p>	<p>Healthcare sector</p> <p>(Certified Nursing Assistant (CNA); Medication Aide; Pharmacy Technician; Phlebotomist; Medical Billing and Coding; Radiology Technologist; Surgical Technician; EKG Technician; Medical Assistant; Licensed Vocational Nurse (LVN); and Registered Nurse (RN))</p>	<p>The target groups for the program include the unemployed, dislocated individuals and incumbent health care workers, as well as Spanish speakers and first generation college students. Program participants must have a high school diploma or GED to enroll.</p>	<p>North Central Texas College (NCTC) trains participants from targeted groups for careers in the healthcare sector, with grant resources providing funding for scholarships for students to attend these programs that they otherwise may have difficulty attending due to financial constraints. Participants can complete basic skills training, as well as training in health care career pathways. The program has a career ladder/lattice approach with an emphasis on “open entry/open exit” modules. This model allows for access to short-term training to pursue additional credentials while working. Credit and non-credit programs are offered to program participants. Non-credit programs are approximately three months in duration. An externship is also required for some programs. Credit programs range in duration from one to two years. Support services are available and include individual assessments, customized remediation options, integrated job readiness skills, as well as basic support services such as child care. Workforce readiness skills training and job search training are conducted during class time. Career advising services are provided by intake and occupational advisors.</p>

Section 3 – Developing and Implementing Random Assignment for the Evaluation

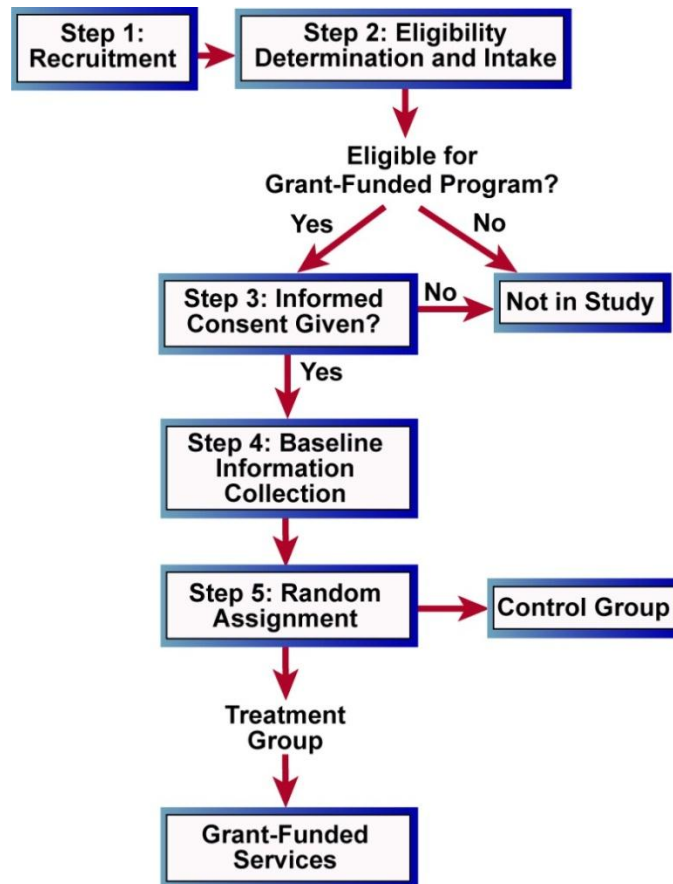
The evaluation will use an experimental design to determine program impacts on participants. In each site, random assignment splits the pool of eligible applicants into two groups: a treatment group that participates in the grant programs under study and a control group that does not. Random assignment ensures that the two groups are as similar as possible—on both measured and unmeasured characteristics—and that any subsequent differences in outcomes between the two groups can be attributed to the grant-funded program services.

This section of the report describes how the evaluation team established the random assignment process in each site. Specifically, the chapter considers collecting baseline data for individuals in the study, and monitoring the integrity of the design over time.

3.1 The Random Assignment Process

Each site has its own plan for conducting random assignment, based on a generic approach that all sites share. Figure 3-1 depicts the generic approach. Application of this approach in each site has been tailored to take into account local practices for recruitment, eligibility determination, and enrollment.

Figure 3-1. Random Assignment Process



As Figure 3-1 shows, the key steps are as follows:

Step 1: Recruitment. Program staff recruits potential participants using their established methods, which can include referrals from community partners, word of mouth, and publicizing service availability through the media.

Step 2: Eligibility. Program staff determines eligibility for the grant-funded services following standard procedures.

Step 3: Informed consent. Program staff informs eligible individuals about the study using a short information sheet describing the study. Staff uses a script, developed by the evaluation team, to describe the purpose of the form. Staff then administers the informed consent form, which describes the study and requires individual to sign the form if they wish to participate in the evaluation. Those who refuse to sign the informed consent form are not be included in the study and are not be eligible for the grant-funded services. They receive information about other services in the community. Because all applicants have sought out grant-funded services, we do not expect many to opt out of the evaluation in this way.

Step 4: Baseline data. Eligible individuals who consent to be in the study complete the baseline information form (BIF). This information is collected in addition to any other intake information the staff collects. Program staff enters information from the BIF into a web-based Participant Tracking System (PTS) developed specifically for the evaluation.

Step 5: Random assignment. Following completion of the BIF, site staff uses the PTS to randomly assign individuals to the treatment or control group. Staff first enters items deemed essential for random assignment to occur (name, date of birth, and Social Security Number) and indicate that the respondent has completed the paper BIF and has signed the informed consent form before clicking the “randomly assign” button. The results of the random assignment are immediately returned to the PTS user, who relays the result to the applicant. If an individual re-applies for the program at a later time, the PTS will identify that individual as already having gone through random assignment and the individual will continue to be treated according to the original assignment (see Step 6 below).

Step 6: Services according to random assignment status. Following random assignment, those assigned to the treatment group are offered the training provided through the grant-funded program while those assigned to the control group are not able to participate in the grant-funded program but can access other services in the community. As noted further below, a key role of the evaluation team is assisting the sites to ensure that those assigned to the control group do not access grant-funded services.

3.1.1 Developing Site-Specific Processes

While Figure 3-1 depicts a general random assignment process, the evaluation team has tailored procedures to fit local conditions. A two-person evaluation team worked with each site to establish mutually agreed upon procedures for conducting random assignment. A general principle was that sites should change as few of their usual procedures as possible for the study. Retaining as many standard procedures as possible helps ensure that the study’s findings are generalizable to grant operations as they would operate in the absence of the study. Specific areas of focus in developing these procedures included:

Participant flow. For the study, each site will recruit and randomly assign from 900 to 1,200 individuals over a 12 to 18 month period. To ensure this is feasible, the team will assess the number of individuals recruited for each training session, and the number expected to be recruited over the study period. All sites will need to increase recruitment to ensure a sufficient sample for random assignment to the

treatment and control groups. The evaluation team will discuss with the grantee any steps it will take to ensure an adequate flow of applicants for the evaluation.

Eligibility determination and enrollment. The evaluation team determined how to incorporate the evaluation-related data collection and random assignment procedures into the site’s intake process. Regardless of whether intake occurs one-on-one or in groups, the team addressed whether intake is centralized or occurs in multiple locations; what information is collected and through which medium (i.e., via paper or input directly into a PTS); who determines eligibility and how; and how much time elapses between applying for program services and starting services.

Point of random assignment. A key principle is that random assignment should occur as close to the beginning of program operations as possible. Because data are collected and analyzed for all individuals in the sample regardless of whether or not they receive services, the team worked with each site to identify a point of random assignment that minimizes the time between random assignment and service provision. This was intended to minimize “no shows”—those assigned to the treatment group who never receive the grant-funded services.

Staffing. The evaluation team will examine staffing arrangements and responsibilities for recruitment and intake in the grant-funded program, in order to determine which site staff members should be involved in evaluation-related data collection and random assignment.

Using this information, the team develops a flow chart mapping the process from recruitment through random assignment that is as minimally disruptive to the sites as possible.

3.1.2 Site-Specific Manuals

For each site, the evaluation team developed a customized manual, with DOL review, regarding site-specific random assignment procedures. The manuals guide site staff through the random assignment process in their local context. These manuals help ensure that staff both implement random assignment procedures accurately and maintain the integrity of the experimental design. The manuals also help to minimize burden and disruption over the course of the study since the manuals serve as references throughout implementation.

The manuals are a one-stop source of information for site staff, on the following topics:

- An overview and rationale for the study and use of an experimental design, such as why the study was funded, what its goals are, and how an experimental design meets those goals.
- The site-specific flow of activities, including the point of random assignment. This part of the manual uses a flow chart to depict the steps from recruitment to random assignment, as well as what happens after random assignment. The manual also includes written instructions for each step, including which staff is involved.
- Procedures for seeking informed consent from potential study participants, including scripts for staff to explain the study, data collection, and the informed consent process.
- Procedures for administering the BIF and entering its contents into the PTS.
- Procedures for conducting random assignment using the PTS.
- Resources for addressing problems that may arise during the study’s implementation. These resources include a list of frequently asked questions regarding random assignment; periodic

check-in phone calls (see below); and a toll-free telephone line that program staff can use to contact evaluation staff.

- Copies of all relevant forms, including the informed consent form and the BIF.
- With site feedback on the manuals, the evaluation site team trained staff on procedures, and each site staff person received a copy of the manual.

3.1.3 Site Training

Comprehensive training of site staff on random assignment procedures is critical to proper implementation of the evaluation. As with the procedures manual, the training sessions were tailored to each site but followed a general approach. The training sessions included an overview of the study, a detailed description of each step in the random assignment process, and hands-on instruction on how to use the PTS. Other topics included data security, protection of human subjects, informed consent scripts, how to prevent and address violations of the research protocols, and additional supports available to site staff. Site staff responsible for conducting study intake and using the random assignment system participated in all aspects of the training, while other site staff (such as managers or supervisors) attended some or all of the training, as needed. The training occurred immediately before the first day of random assignment, after which the evaluation team remained on-site to observe random assignment and provide technical assistance. Some areas the team monitored include:

- Is staff describing the study clearly?
- Is staff responding to questions from potential participants appropriately?
- Are potential participants expressing concerns about the study or random assignment? Can scripts or training materials be adjusted accordingly?
- Is the web-based PTS working smoothly? Is staff having difficulties entering data and randomly assigning participants?

Evaluation team members will conduct refresher or new staff trainings via webinar as needed.

3.2 Baseline Data Collection

This section describes key forms that are collected during the random assignment process and data quality control measures. Staff training included instructions on how to administer the forms and return them to the evaluation team.

3.2.1 Participant Consent Form (PCF)

Study participants' provision of informed consent is a critical part of the intake process. The informed consent process includes a review of what participation in the study entails, and a collection of the signed Participant Consent Form. As required by federal regulations and industry standards regarding the protection of human subjects engaged in research projects, not only does the individual give consent to participate in the study, but also indicates his or her understanding of what study participation means. For the GJ-HC Impact Evaluation, this includes consenting to be randomly assigned to either the treatment or control group and to remain in the assigned group for the duration of the study; agreeing to participate in initial and follow-up data collection; an understanding that he or she will receive an incentive for follow-up surveys; and allowing the study team to collect individual-level administrative data (e.g., UI wage records).

Program staff explain the informed consent form to eligible applicants, distribute it, and ask individuals to read and sign the form if they consent to participate in the study. Those who consent sign the form and return it to the staff person. On a regular schedule, program staff batch and return signed PCFs to the evaluation team via Federal Express.

3.2.2 Baseline Information Form (BIF)

We discuss the content of the BIF in Section 4. Here we address how and why we incorporate the BIF into program processes. Program staff administers the BIF to those who sign the informed consent form. The BIF collects demographic and socioeconomic characteristics, and employment history. It also collects critical contact information for the individual and three family members or friends who can help locate the individual during the follow-up period. Data from the BIF are needed for multiple purposes, including:

- Conducting random assignment (basic identifying information is needed prior to conducting random assignment)
- Monitoring random assignment (ensuring that no one goes through the random assignment process more than once)
- Locating participants for surveys and collecting administrative data
- Describing the sample at baseline
- Defining subgroups for the impact analysis
- Increasing the precision of impact estimates
- Adjusting for non-response bias

The evaluation team explored gathering this information from currently available data sources, including the quarterly reports that grantees submit to DOL. The team found that these sources fail to provide key identifiers and contact information for future data collection, to document the characteristics of the sample in terms of education and employment history and work-related barriers, and to supply key data needed to create meaningful subgroups.

As with enrollment and random assignment procedures, the process for administering the BIF and entering data into the PTS is tailored to each site.¹ The typical scenario is that participants fill in a paper form and return it to the site staff. Staff reviews the form to ensure it is complete and legible, seek clarification as needed, and then enter the fields into the web-based PTS. The PTS includes a check box indicating that the BIF was completed. This box needs to be checked before an individual can be randomly assigned. As with the PCF, the BIFs are batched and returned to the evaluation team at regular intervals.

3.2.3 Data Quality Control

It is important to ensure that the study participants are enrolled in a timely manner and that the data collected are of high quality. For the former, the sites need to complete enrollment of participants within a

¹ Data provided by grantees in DOL's RAD system will not be used by the evaluation due to its limited overlap with the variables needed for the evaluation, the need to ensure complete consistency of data collection procedures for the baseline surveys (just as will be done on the study's follow-up surveys), and ease of data access (i.e., data transfers out of RADs into the evaluation PTS will not be needed).

time frame that allows for the follow-up data collection, analysis, and reporting phases of the study to occur within the contract period for the evaluation, as detailed below in Section 7. With regard to the latter, complete contact information is critical for locating sample members for the purposes of follow-up data collection.

Participant Tracking System (PTS)

The evaluation team will monitor sample build-up and data quality through careful review of the data in the PTS and frequent communication with the sites. The PTS includes default checks to ensure accuracy of information. Social Security Numbers are entered twice and checked for match. Additionally, the PTS performs a background check on the age of the applicant and flags any dates of birth that indicate that the applicant is either under 18 years of age or over 65. While these age groups are allowed, they would be rare. Instead, these entries are likely to be errors. Thus, the PTS is designed to alert the user to these entries. The PTS also checks for duplicate entries of applicants (both within and across sites) via SSNs and via name and date of birth. If an entry appears to be a duplicate, the site staff is not allowed to enter that applicant until contacting the site's liaison at Abt or Mathematica.

The evaluation team has access to all data entered into the PTS and is able to monitor sample build-up on a real-time basis. A slowdown in the pace of sample buildup from the projected rate might signal a recruitment problem that will need to be discussed with the grantee and with appropriate staff at DOL so that remedial steps can be taken. Furthermore, an imbalance in the size of the research groups, compared to the planned distribution of sample into the two research groups, could indicate a possible problem with random assignment.

Data entered into the PTS is also checked for quality throughout the random assignment timeframe. If a site consistently enters only limited data, the evaluation team will work with that site to improve data entry. The evaluation team staff will also do random checks of hard copy BIF forms to ensure that site staff is accurately and completely entering data into the PTS. If the data are not complete we will investigate the causes and work with site staff to obtain the omitted information. In addition, we will check the PTS file regularly to confirm that no important differences in baseline characteristics have arisen between the treatment and control groups.

As part of our use of the PTS, we will send regular progress reports regarding random assignment to the sites and to DOL (monthly, or more frequently if problems are identified) and discuss any issues and corrective actions required.

Ongoing Communication with Sites

The evaluation team will conduct periodic check-in phone calls with sites where the designated evaluation site liaisons support the sites and monitor how implementation of the study is progressing. Because of the likelihood of issues arising during the early phase of the random assignment and data collection period, these phone calls occur weekly during the first several months of random assignment and, at a minimum, monthly thereafter. The calls involve the key site staff, including the primary evaluation liaison, and others such as supervisors or intake workers as needed. Based on the information that emerges during these calls, the evaluation team will determine the appropriate strategies to address any issues. The strategies could include clarifications about study procedures or other types of technical assistance. As needed, the site liaisons will draw upon other evaluation team members to ensure that strategies that are used and decisions that are made are consistent across sites.

3.3 Randomization Procedures

Earlier parts of this section addressed the development of random assignment procedures, training staff on random assignment and data collection, and monitoring of random assignment and data entry. This section describes how random assignment is operationalized. It begins with the properties of random assignment, then discusses stratification and blocking, and concludes with a description of how the PTS operationalizes random assignment.

3.3.1 Goals and Properties of Random Assignment

As noted above, a carefully maintained random assignment process ensures that the only systematic difference between treatment and control group members is that the treatment group members have access to grant-funded services while control group members do not. Importantly, members of both groups have access to all services available to them in the community that are not funded by the grant. In this way, the control group’s experiences can provide information about what would have happened in the absence of the grant (and the study); this situation is called the counterfactual. Differences in the outcomes of the two groups—the impact estimates—can then be attributed to the difference in access to the grant services. Randomization also ensures that subsets of the treatment and control groups defined by the baseline characteristics are also statistically equivalent, making outcome comparisons between treatment and control group members in any subpopulation unbiased.

The equivalency of the two research groups holds regardless of the *share* of eligible applicants placed in the treatment group. That is, the treatment group can comprise any percentage of the total sample and their characteristics should still be identical to the control group on all measured and unmeasured baseline characteristics except for chance deviations. While the fraction assigned to the treatment group, or equivalently the random assignment ratio (share assigned to the treatment group divided by the share assigned to the control group) affects our ability to detect non-zero impacts as statistically significant (discussed later in the report), it does not alter the absence of bias among treatment-control comparisons as measures of program impacts. Indeed, the random assignment ratio can differ across sites—assuming there are important non-statistical reasons for the ratio chosen in each individual site—and can even change over time within a given site (for example, starting out as a 50-50 treatment-control ratio and shifting to a 70-30 ratio). Nothing about the propensity for random assignment to ensure unbiased impact estimates is threatened by “unbalanced” designs of these types. Thus, where it proved essential for insuring grantee participation in the study, alternative random assignment ratios were adopted in some sites (see Figure 2-2 above).

3.3.2 Stratification, Pre-Designation, and Block Size

For any given sample size and random assignment ratio, the results of randomization can be improved through stratification, pre-designation, and blocking. Each of these elements is described below, though they also intersect in important ways.

Stratification, or selecting the sample to ensure that individuals of a given type are split between the treatment and control groups in the same ratio as the overall sample, can help improve the results from random assignment by making the overall treatment group and control group more alike. For example, if the overall treatment-control ratio were 70-30, stratification on sex would assure that 70 percent of all women randomized are assigned to the treatment group, as are 70 percent of all men. If stratification were not used to control these percentages, small deviations would be possible; for example, by chance 66 percent of women and 74 percent of men might be assigned to the treatment group, making the treatment group more than half male and the control group more than half female, making the two groups less alike.

These would be *chance* differences and would not cause selection bias in the impact estimates, but they would make it harder to detect a non-zero impact in an effective program.

To implement stratification when individuals are randomized one at a time, as in the current evaluation, it is necessary to apply a second refinement to random assignment: pre-designation. Pre-designation of slots to treatment and control group involves placing an eligible applicant in the next “unclaimed” slot on a list when his or her time of random assignment arrives. By setting up the list in advance, the shares of treatment and control group members assigned can be directly controlled. For example, in the case of a 70-30 random assignment ratio, the list can be set up so that over the course of N assignments exactly $0.7N$ applicants are assigned to the treatment group and exactly $0.3N$ to the control group. This contrasts with conventional random assignment in which each person is independently randomized with a 0.70 chance of entering the treatment group and 0.30 chance of entering the control group; in such a system, it is possible for the independent chance outcomes of randomization to designate, say, $0.63N$ applicants as treatment group members and $0.37N$ as control group members. Randomness is preserved in the pre-designated-list approach by randomly ordering each set of $0.7N$ treatment group slots and $0.3N$ control group slots.

The use of pre-designated lists of treatment and control group slots of this sort allows the team to stratify the random assignment to ensure a treatment-control balance for subtypes of individuals, such as women and men. This is accomplished by the simple expedient of creating *separate* lists for each subtype of individuals, and placing each individual to be randomized into the next open slot *for that subtype*. This ensures in the example above that the exact $0.7N$ treatment group assignments and exact $0.3N$ control group assignments among N total assignments hold for each subtype—i.e., for women and for men.

The most common stratifier used in random assignment experiments—and one we use here—is study site. A separate pre-designated list of treatment and control group slots was created for each of the four study sites, to make sure that site-specific impact analyses are based on samples that tightly match the desired treatment-control ratio in each site. Such stratification is also obviously necessary if different ratios are desired in different sites. Because parsimony of stratification is important (Bloom, 2005), the team did not create separate pre-designated lists of treatment and control group slots for different subtypes of individuals *within* a site.

The final aspect of the randomization approach to be specified is the “block size”—i.e., the value of N for which, in the above example, $0.7N$ slots are designated as treatment group slots and $0.3N$ slots are designated as control group slots. Blocks, or sequences of randomly ordered treatment and control group slots, cannot be too long if they are meant to effectively control the treatment-control ratio achieved; for example, a block size of 10,000, when the target total sample size is 1,200, would do little to ensure balance when randomly assigning those 1,200 individuals in a given site. While this is extreme, the principle applies more generally: smaller blocks impose tighter control on the achieved treatment-control ratio. At the same time, the team does not want to use very small blocks lest the site staff conducting random assignment begin to anticipate the outcome for the next individual submitted. For example, in a 50-50 design if $N = 2$ —and site staff know or begin to suspect this—any time two consecutive assignments go into the control group staff will know that the next person randomized in their site will enter the treatment group, since two consecutive control group assignments must come from adjacent blocks in a sequence T-C / C-T, where the slash mark designates the end of one block and the beginning of another. Even blocks with $N = 4$ might be readily recognized by site staff, allowing them to ensure that preferred applicants get into the treatment group or at least increase the odds that this occurs by choosing the order in which cases are randomized. As soon as specific types of individuals gain a higher or lower

chance of being in the treatment group than other individuals, “gaming” occurs, randomization is broken, and selection bias produced by the selecting being done by site staff creeps into the impact estimates.

To balance control over the treatment-control ratio in each site with protection from random assignment “gaming”, the team decided to:

- Use blocks of size 4 and of size 6², and
- Randomly sequence the blocks themselves along with the slots within each block

This assumes a 50-50 random assignment ratio applies in each site, so that any even number can be used as a block size. This design effectively ensures that site staff will not figure out the blocking structure and hence the random assignment sequence on the list. Staff were not told that pre-designated lists exist, that they are organized into blocks, or that the size of these blocks varies between four and six slots. Simultaneously, it precludes the ratio of treatment and control group assignments from moving very far away from the exact desired ratio at any point in time. The latter is important statistically as discussed above, and operationally to make sure that no grantee hits a “bad luck streak” during random assignment during which seven or eight consecutive applicants are randomized into the control group.

3.3.3 Randomization through a Computer Platform

The PTS is used to randomly assign applicants to either the treatment or control group. After staff completes the items deemed essential for random assignment to occur (name, date of birth, and Social Security Number) and indicates that the respondent has completed the BIF and signed the PCF, the site staff will click on a “randomly assign” button. The results of the random assignment will be immediately returned to the PTS user, who relays the result to the applicant. Note that, depending on operations at each site and the needs/preferences of the site staff, our approach allows applicants to either be randomly assigned one at a time or to be randomly assigned in a group, (or sometimes one at a time and other times in a group).

The internal computer process that occurs when the “randomly assign” button is clicked works as follows. Embedded in the PTS is a random assignment table containing the 1,600 slots for each site (an extra 400 to 700 slots were included in the event any site exceeds its goal of 900 to 1,200 applicants). Each slot has a treatment (T) or a control (C) assignment. Applicants are placed into the next-available slot in the table and thus assigned to the T or C group. Figure 3-2 below provides a sample of one of the random assignment tables for a site. Note that each site has its own random assignment table.

The random assignment table was generated with SAS 9.2, using the PROC PLAN procedure. This procedure randomly assigns slots to a T or C status based upon user input that indicates the probability of selection, the number of slots to be generated, the number of blocks within which slots will be grouped, and the size(s) of those blocks (which varies between the four- and six-slot blocks described earlier). Within each block, half of the slots are Ts and half are Cs. The blocks are randomly intermixed throughout the random assignment table. Figure 3-3 below provides a sample of the blocking of slots and the random order of block sizes throughout the table.

² These block sizes will be appropriate for sites using a 50-50 random assignment ratio. For sites with different ratios, different block sizes will be needed. For example, the 67-33 ratio in Grand Rapids necessitates, block sizes that are multiples of 3 (e.g., blocks of size 6 and 9). If the random assignment ratio *changes* in midstream in a given site, all unused blocks will be deleted and replaced with new blocks that reflect the new ratio.

Figure 3-2. Example of a Random Assignment Table and Slots

RecordID	Assignment	Probability	Site_Number	BlockID	Used
10001	Treatment	0.5	1	1	X
10002	Control	0.5	1	1	X
10003	Treatment	0.5	1	1	X
10004	Control	0.5	1	1	X
10005	Treatment	0.5	1	2	X
10006	Control	0.5	1	2	
10007	Control	0.5	1	2	
10008	Treatment	0.5	1	2	
10009	Control	0.5	1	3	
10010	Control	0.5	1	3	
10011	Control	0.5	1	3	
10012	Treatment	0.5	1	3	
10013	Treatment	0.5	1	3	
10014	Treatment	0.5	1	3	
10015	Treatment	0.5	1	4	
10016	Control	0.5	1	4	
10017	Treatment	0.5	1	4	
10018	Control	0.5	1	4	

The next applicant to be randomly assigned will be placed into this “slot,” and thus placed into the control group. Slot assignments of “Trt” (treatment) and “Ctr” (control) were randomly generated with SAS.

Figure 3-3. Example of Blocking within the Random Assignment Tables

RecordID	Assignment	Probability	Site_Number	BlockID	Used
10001	Treatment	0.5	1	1	X
10002	Control	0.5	1	1	X
10003	Treatment	0.5	1	1	X
10004	Control	0.5	1	1	X
10005	Treatment	0.5	1	2	X
10006	Control	0.5	1	2	
10007	Control	0.5	1	2	
10008	Treatment	0.5	1	2	
10009	Control	0.5	1	3	
10010	Control	0.5	1	3	
10011	Control	0.5	1	3	
10012	Treatment	0.5	1	3	
10013	Treatment	0.5	1	3	
10014	Treatment	0.5	1	3	
10015	Treatment	0.5	1	4	
10016	Control	0.5	1	4	
10017	Treatment	0.5	1	4	
10018	Control	0.5	1	4	

The first “block” in this random assignment table contains 4 slots, among which 2 are treatment and 2 are control.

The next “block” in this random assignment table also contains 4 slots, among which 2 are treatment and 2 are control.

The third “block” in this random assignment table contains 6 slots, among which 3 are treatment and 3 are control. Random ordering of block sizes continues until 1,600 slots are generated.

Section 4 – Data Collection

To support an evaluation of the scope the Green Jobs and Health Care Impact Evaluation, we will draw on several sources of data. This section describes the source of primary and extant data for the evaluation. First we detail the primary data, which comes from baseline, 18- and 36-month surveys. Then we explain sources and uses of administrative data for employment and earnings. Finally, we review the data to be collected for the process study.

4.1 Survey Data

Data collected through three surveys—a baseline information form to be completed before random assignment, a survey completed 18 months after random assignment, and another survey completed 36 months after random assignment—will be used to facilitate and enhance the impact analysis in several ways. Below we describe each of the instruments and how the data collected from each will be used.

4.1.1 Baseline

The information collected on the BIF serves several purposes, as noted in Section 3.2.2. Figure 4-1 lists all of the elements to be collected at baseline.

4.1.2 Follow-up Surveys

All study participants, including both treatment and control group members, will complete the first survey 18 months after their random assignment dates and the second survey 36 months after random assignment. The follow-up surveys serve two purposes. The first is to collect information on service receipt and educational outcomes, and the second is to examine long-run employment, and economic security (see Figure 4-2). While each wave of the survey addresses both issues to some extent, given their timing in relation to individuals' participation in training, the first will focus on service receipt and educational attainment and the second on employment, earnings, and career progression.

Data on service receipt, a focus of the 18 month survey but also addressed in the later survey, will aid in developing an understanding of any program impacts. Control group members may have access to similar types of training services provided or funded by sources other than the grantee; this could result in smaller differences in outcomes between treatment and control groups, reducing the magnitude of impact estimates. Without knowing what services were received by control group members, it is possible that a small or null impact could mistakenly be interpreted as evidence of a lack of program effectiveness. It is therefore important to have an understanding of the services received by both groups in order to interpret the impact estimates.³

Additional data elements from the 18-month survey will serve as short-term outcomes in the interim exploratory impact analyses. In particular, the survey will collect information on short-term outcomes of interest in domains such as the acquisition of credentials; employment-related outcomes aimed at capturing the quality of the job and the match of the type of job with the training program. It will also

³ Tests measuring specific knowledge or skills gained through the training programs. are not feasible to conduct under the scope of this evaluation. Instead, we will measure the curriculum and topics covered by survey-reported training programs, which, together with completion rates, will provide a reasonable proxy for what participants learned.

collect data on public assistance benefits receipt and income, as well as opinions about work parallel to those collected on the BIF to determine the extent to which they vary by treatment status.

The 36-month follow-up survey will collect similar data but focus on long-run employment and earning as well as snapshots about a study participant’s status at the time the survey is administered. This survey will document employment and income since the time of random assignment, wage and earnings progression, career advancement, other issues related to job quality (such as employee benefits), and use of other public benefit programs.

Figure 4-1. Baseline Data Elements

Personal and Contact Information
Full name
Social Security Number
Date of birth
Address
Telephone numbers
Email address; Facebook name
Contact Information for Friends and Relatives
Addresses, telephone numbers, email for up to three friends or relatives
Demographic and Socioeconomic Characteristics
Highest level of education completed
Currently enrolled in school or training
Previous participation in education and training
Sex
Ethnicity
Race
Primary spoken language
Number of children
Age of youngest child
U.S. citizenship status
Ex-offender status
Disability status
Employment

Currently employed?
Number of hours per week
Wage
Employed in last 12 months?
Number of months worked
Annual earnings
Public Assistance Receipt and Housing Status
Receipt of TANF, SNAP
Receipt of UI benefits
Weekly benefit amount
Section 8 or public housing assistance
Own or rent home
Opinions About Work
Factors that limit ability to work
Job preferences and motivation
Lowest acceptable hourly wage

Notes: TANF = Temporary Assistance for Needy Families; SNAP = Supplemental Nutrition Assistance Program; UI = Unemployment Insurance

Figure 4-2. 18- and 36-month Survey Data Elements

Service Receipt	
Employment-related	Assessments/educational achievement
	Unpaid work experience
	Case-management/counseling
	Follow-up/retention services
Training/Education	Type of basic education course
	Secondary or post-secondary education
	Occupational skills training
	Occupation for which being trained
	Duration and hours attended
	Program completed or not/why not

	Highest math course taken in high school ^a
Supportive and other services	Type of supportive services
	Received needs-related payments
	Other services received
Perspectives on services, if received any ^a	How heard about training/employment services
	Why chose to seek training/employment services
	Reasons for non-use of services (as applicable)
Educational Outcomes	
Acquisition of credentials	Completion of training/education
	Attained a degree, license, certification, or other credential
	Type of degree, license, certification, credential
	Received a high school diploma or GED
	Where obtained degree, diploma, license, certification, credential
Employment and Earnings	
Employment since program completion	Employed
	Earnings
	Wage rate and hours worked
	Industry/Occupation
	Industry/occupation aligned with training?
	Length of time in current job
	Availability of fringe benefits
	Job is temporary/permanent
	Job is on a career pathway
	Job is unionized
	Number of jobs held
	Industry/occupation of previous jobs

Figure 4-2. 18- and 36-month Survey Data Elements (Continued)

Employment and Earnings	
Employment since program completion	Previous jobs aligned with training?
	Number of and reasons for job separation
Income and use of public benefits outcomes	Receipt of TANF, SNAP, SSI, Medicaid ^b , or other benefits
	Receipt of UI and Trade Adjustment Assistance benefits
	Household income and poverty status
Other	
Opinions and work preferences	Factors that limit ability to work
	Job preferences and motivation
	Lowest acceptable hourly wage
Criminal behavior ^c	arrests, incarceration
Financial stability	hardship, debt, assets
^a Items will be asked only on the 18-month survey. ^b This item will measure only program participation status (yes/no), not receipt of specific Medicaid-funded health care service. ^c This item will be asked only in Grand Rapids, the site where an important share of participants are expected to be ex-offenders.	

4.1.3 Survey Data Collection Procedures

Assuming a response rate of 80 percent, we anticipate approximately 3,200 completed surveys for each round of survey data collection.

Achieving high response rates. We will use a combination of telephone data collection and field locating to ensure the highest possible response rates. Subject to Office of Management Budget (OMB) approval, we will offer sample members a \$25 incentive for completing each follow-up interview. The incentive encourages sample members to participate in the survey and to provide updated contact information, especially during the 18 months between the first and second follow-up surveys. Through this combination of telephone and field locating resources, along with the proposed \$25 incentive, we expect to achieve at least an 80 percent response rate for each survey. Since we anticipate contacting all 4,024 study participants,⁴ which results in a target of about 3,200 completed surveys—a target that past experience in similar evaluations says is achievable.

Achieving high response rates to baseline and follow-up surveys will require a combination of techniques that Mathematica has refined over the past 40-years, including:

⁴ Current estimates, shown in Figures 2-2 and 6-1 above, predict 4,024 combined treatment and control group members across the four sites.

- Compelling advance materials, including brochures about the study, FAQs, and endorsements from leading organizations;
- Assurance to sample members that the information they provide will be secure, treated confidentially to the extent permitted by law, and used only for research purposes;
- Well-designed questionnaires, with cognitively tested and easy-to-answer questions;
- A toll-free help line for sample members to call with concerns or to schedule an appointment and well-trained interviewers able to address sample member's concerns;
- Multiple attempts to reach respondents at various times of the day and week;
- Specialized refusal conversion and training as needed; and
- Providing a monetary thank-you (as determined by OMB) to show appreciation for a participant's time and effort.

Instrument development and pre-tests. To ensure that the data collection methods result in complete and high quality data, the development of the questionnaires will proceed through a systematic process that includes (1) collaborating closely with ETA to ensure that information collected provides complete answers to the research questions; (2) reviewing the existing measures to select valid and reliable scales for inclusion and developing new measures as needed; and (3) conducting a well-designed pretest that will ensure that questions are understood as intended. We will develop the questionnaires in collaboration with ETA and will start by reviewing previous questionnaires, notably from projects related to the Individual Training Account program, the Trade Adjustment Assistance Act, and Project GATE as these questionnaires capture similar data and their questions have been approved by the OMB.

Prior to OMB approval, we will pretest the questionnaires with nine or fewer program participants from one site who are not members of the research sample. To obtain accurate timing estimates, we will want to mimic actual field conditions; therefore we will conduct the pretest over the telephone, reading the questions verbatim and following the order as established in the questionnaire. After the questionnaire is completed and timing estimate obtained, interviewers will then spend a few minutes debriefing the respondent to check on question comprehension and clarity as well as confidence in answers. We will use multiple interviewers for the pretest to ensure that our observations are not attributable to the interviewer rather than the questionnaire itself. (For the full study, our comprehensive training addresses avoiding interviewer bias.) As the interviewer conducts each interview by telephone, additional staff members will observe the interview and take detailed notes (including timing estimates). After an interview is completed, interviewers summarize their findings, question-by-question.

Because items appearing later in a questionnaire may influence a respondent's understanding of an earlier item, we may not be able to accurately assess comprehension of early items unless we also incorporate concurrent probing into the pretest; however, if we do, then our timing estimates will be invalid. Therefore, to obtain timing estimates, half of the pretests will be conducted as described above, and the remaining half will incorporate cognitive interviewing techniques (e.g., concurrent probing and think-aloud protocols) to assess respondents' comprehension and clarity. A survey researcher trained in cognitive interviewing techniques will conduct the cognitive interviews, digitally recording the session in order to accurately summarize the findings.

A pretest report will identify the problematic questions and interviewer training issues as well as make recommendations for improvement. ETA staff members can participate in the pretest process by

monitoring remotely in real time, or requesting that we digitally record the interviews for their future review.

Once ETA approves the final questionnaire, we will prepare a detailed CATI specification document for the Blaise programmers, who will use the document to program the instrument into the Blaise software. The specification's document is organized on a question-by-question basis and includes question routing, hard and soft range checks, and any constructed variables referenced in the questionnaire.

Once the instrument is programmed, we will test it from multiple perspectives—first from CATI programmers and then the survey project staff—all following a pre-specified testing plan created for the project. The testing plan maps out all of the critical items and paths in an instrument; it also provides the framework for documenting testing progress and specific scenarios the testers should follow while testing the instrument. Typically, the testing done by the CATI programmers ensures that the instrument is free of obvious errors (such as missing questions or answer categories) and then verifies logic skips, text fills, edit checks, displays, and adherence to Blaise screen standards and guidelines. After the CATI programmers have completed their testing and updated the instrument, the survey person most familiar with the CATI specifications will test the instrument carefully, identifying and documenting any serious design flaws or unexpected changes in the layout of the instrument. After the CATI programmer makes changes based on this feedback and both sides agree on the structure and layout of the instrument, a new version of the CATI instrument is released and testing continues. At this point the assembled team of testers (which can include ETA, if so desired) takes over and follows the testing plan. When they have tested the instrument thoroughly and revisions have been made, staff at Mathematica's Survey Operation Center (SOC) will test the instrument, looking for issues specific to administering the questionnaire.

Interviewer training. Senior project staff and professional staff from the SOC will train the interviewers. We will train 50 telephone interviewers for this project and anticipate that most of the interviewers assigned to this project will be experienced CATI interviewers and will already be thoroughly familiar with standard interviewing techniques and with specific SOC procedures. However, if it proves necessary to add additional resources, the new staff, like all new interviewers at the SOC, will receive generalized training on interviewing techniques.

In any event, all interviewers, including new and experienced, who are assigned to the project will receive project specific training. This will include classroom and self-guided instruction. Prior to the classroom training, interviewers will receive a customized CATI training manual, that includes an overview of the project, frequently asked questions, and question-by-question objectives for each item in the survey instrument. Interviewers will be given two hours to review in advance and prepare for an in-class quiz on the materials.

During classroom time the trainees will be guided through the instrument with emphasis on unique or especially challenging aspects of the follow-up questionnaire, with suggestions for responding to respondent questions about the survey. Classroom training will also include "refusal avoidance training" where audio recordings of exemplary interviewer-respondent dialogue will be played which focus on ways to avoid refusals as well as methods to convert a refusal once it occurs. Paired interviewing practice will also occur with supervisors monitoring to offer suggestions along with appropriate probes that are specific to the situation.

Supervisors monitor the telephone interviewers frequently to ensure that we achieve high data quality and consistency. They pay special attention to difficult items such as questions about employment history, occupation and industry, and specific education and job training programs where responses might be inconsistent or inadequate.

The telephone monitoring system allows the SOC supervisors, monitors, and clients to listen to interviews without either the interviewer's or the respondent's knowledge. It also allows the supervisor to view an interviewer's screen while an interview is in progress. Interviewers are informed that they will be monitored, but they do not know when observations take place. Supervisors concentrate on identifying behavioral problems involving inaccurate presentation of information about the study; errors in reading questions; biased probes; inappropriate use of feedback in responding to questions; and any other unacceptable behavior, such as interrupting the respondent or offering an opinion about specific questions or the survey. Supervisors will review results with the interviewer after each monitored interview or upon the change of shift. Results of monitoring are maintained electronically, and supervisors review performance on specific calls or interviews, as well as evaluate interviewers' progress.

Survey administration. Professional telephone interviewers from Mathematica's Survey Operation Center (SOC) will administer the survey by telephone with Blaise computer assisted telephone interviewing (CATI) software. The software maximizes data quality by enforcing question skip logic and by checking data items as they are entered to make sure they are in appropriate ranges and are consistent with previous responses. CATI call scheduling, including time zone adjustments, interviewer assignments, and call priority schemes will be used to increase the efficiency of contacting sample members.

Interviewing progress will be monitored on a daily basis by SOC Supervisors using reports generated from Mathematica's sample management system (SMS). The reports are generated as interviewers enter the status of each call they place through the course of the day. SOC supervisors regularly review cases sent to their "supervisor review" inbox. Each morning, SOC supervisors review the prior day's effort and success, via SMS reports, as well as check on staffing for scheduled interviewing appointments.

Although inability to find a sample member is the most frequent reason for non-response, interviewers will be trained in refusal avoidance and conversion techniques. If a refusal occurs, interviewers will explain the situation in their interviewer notes section of the instrument, which will allow a customized refusal letter to be crafted and sent by priority mail to the respondent's mailing address. The case will be put on hold for seven to fourteen days and then an experienced refusal conversion interviewer will be tasked with re-contacting the sample respondent to complete an interview

We will address inability to find sample members by using a locating effort that progresses from simple, batch-mode steps to more intensive customized efforts for individual participants—the most resource efficient approach for obtaining a desired response rate. This process begins with collecting extensive contact information on the baseline information form prior to random assignment: names; addresses; e-mails; telephone numbers (land and mobile); Social Security Numbers; and contact information for up to three individuals who would know where participants have relocated. We request this information again at the 18-month survey for use in the 36-month follow-up.

Before each follow-up survey, we mail a "pre-notification letter" describing the study, encouraging participation in the telephone interviews, and providing a toll-free number for participants who have moved or changed telephone numbers to call and update their information. We also use automated telephone look-up services to find telephone numbers for the standardized and newly found addresses.

When a new address or telephone number cannot be found, we continue with more resource-intensive telephone locating. Searches are done one-by-one (rather than in batch mode). We use "Accurint" for new addresses and reverse directories. We check Phones Plus for cell-phone numbers. We call the contact person named at the baseline interview who does not live in the household. We explore Google and social

networking sites for clues. If appropriate, we move on to professional and organizational directories, alumni sites, prison searches and death records.

Despite multiple in-house locating and contact attempts by telephone, however, we have budgeted based on past experience (cited above) that one-third of the sample for the 18-month survey and 40 percent of the sample for the 36-month survey will require locating by field staff that search out the sample member “on the ground.” Specifically, field locators will find sample members through efforts such as talking to neighbors, relatives, postal employees and even local merchants. When a locator finds a sample member, he or she will explain the study and provide a cell phone on which the sample member can immediately dial into the SOC to complete the interview. Limiting the function of field staff to locating, and avoiding in-person interviewing, eliminates the major cost of equipping field staff with laptop computers and training them to conduct the interview. It also avoids the risk of introducing mode effects that we might have if we were to conduct some interviews in person in the field. The field locators will be able to hand a sample member an incentive check as soon as the interview is completed, which is likely to increase cooperation.

4.2 Administrative Data on Employment and Earnings

Administrative data will serve as another source of data on earnings and employment status for study participants for impact analyses. Administrative data on earnings before random assignment will enable us to increase the precision of impact estimates by using past earnings as a covariate in the earnings impact analyses. These data will also enable the study team to define subgroups for exploratory analyses and can serve as non-survey data to verify results of the overall impact analyses using a different measure of the outcomes of interest.

Below we describe the advantages and drawbacks associated with two sources of administrative data: state-provided quarterly Unemployment Insurance (UI) wage records and the National Directory of New Hires (NDNH) which compiles these records from all states. Ultimately, only one source of administrative data will be collected for the study, and for reasons described below, the study team recommends reliance on the NDNH if agreement can be reached with the U.S. Department of Health and Human Services (DHHS, which compiles the directory) to supply these records.

4.2.1 State UI Wage Records

Each state UI agency’s records contain the quarterly earnings, by employer, of all UI-covered employees in the state; thus both earnings and employment status can be derived from this data source. By law, most employers are subject to a state UI tax and are required to report the earnings of each of their employees on a quarterly basis to the state UI agency. More than 90 percent of all workers are covered by these files (Hotz and Scholz 2009; Kornfeld and Bloom, 1999).

The advantage of state UI data records wage files is that they are fairly uniform across states and over time, making them easy to work with from a data processing standpoint. In addition, the administrative data are not subject to recall error and non-response that can occur with survey data collection. Depending on the level of effort needed to obtain these data and any payments to states to pay for accessing these data, they can also be less expensive than surveys.

However, state UI data do have several drawbacks. First, not all workers are covered, meaning that earnings might be undercounted to some extent. Although 90 percent of all workers are covered by wage records, the 10 percent of all workers that is not covered might contain a disproportionate percentage of study participants. This is because workers excluded from UI earnings records include, among others,

federal employees, self-employed workers, workers in service for relatives, some domestic service workers, and some workers who are casually employed “not in the course of the employer’s business” (U.S. Department of Labor, 2004). State UI data also do not cover earnings “off the books” and only cover earnings in the state in which they are collected, which means there will not be records for employees who relocate out of state or who cross state lines for work. Finally, since the wage records must be matched to participants by Social Security Number, there can be errors if the SSN is reported incorrectly by the worker or employer to the state agency or by the worker to grantee staff. However, these issues are mitigated by the fact that they will apply equally to both treatment and control group members, and so will not introduce bias to the analysis.

Another potential drawback to using state UI wage records is that states are not mandated to provide such data to the study and would do so only after entering into a data use agreement with the study team. The evaluation team has had some success obtaining these data for states in the recent past (for example, for the Individual Training Account Experiment, completed in 2006 by Mathematica), although it has been increasingly difficult due to concerns about confidentiality as well as other state-specific issues regarding data access.

4.2.2 National Directory of New Hires

The NDNH, which is administered by the Administration for Children and Families (ACF) in DHHS, contains data similar to the UI wage records. The NDNH is a national database of employment and UI information that is collected to aid states in locating noncustodial parents and enforcing child support orders. State workforce agencies and federal agencies provide information to the NDNH. These data have the same origins as what we would collect from state UI agencies (and the same advantages and drawbacks in terms of which workers and jobs are and are not covered, except that NDNH data also cover federal employment) but has the additional advantages that it is a single file format, covers the entire country, and would require a single data use agreement; this implies that it could be less costly to collect.

The primary potential drawback to the NDNH is access. Some federal evaluation projects have been granted access to NDNH data through an arrangement with ACF, other projects have been denied access. Modest analytic limitations may also arise if this source is used, including retrospective data (at demonstration entry) that goes back only two years and incomplete ability to obtain separate earnings information for all potentially interesting participant subgroups. The study team is currently working with ETA to determine whether the evaluation can obtain earnings information from the NDNH and to better understand any potential implications for the analysis and gauge their importance.

4.2.3 Collecting UI Data

In the end, the study will collect UI wage records either directly from the states of California, Michigan, Minnesota, and Texas, the four states in which the four selected grantees are located, or via the NDNH. If pursuing these data through state agreement, the study will request state UI wage records for each participant for a total of 66 months. The data will be requested in two separate batches—one at 15 months and one at 36 months after the end of the random assignment period. The first extract will cover the 12 months before random assignment began, the 18-month intake period, and the 15 months after the end of random assignment, for a total of 45 months of data. The second extract will cover the subsequent 21 months so that 36 months of administrative follow-up data are available for sample members. Timing the request in this way, rather than requesting all of the data 36 months after random assignment ends, means the data will be more accessible for the states, which usually archive wage records after a few years have

passed.⁵ In addition, the administrative data from the first batch can be analyzed in tandem with data from the 18-month follow-up survey for inclusion in the interim report.

This data collection schedule has implications for the extent to which the administrative data can be used in interim and final impact analyses. The first batch of administrative data collection—15 months after the end of random assignment—will roughly line up with the collection of 18-month survey data for individuals who were randomly assigned late in the intake period. Because there is a lag time of approximately three to six months in the reporting of UI wage records by employers to the states, we may not receive data through the full 15 months after random assignment for these individuals. Similarly, the administrative data requested 36 months after the end of random assignment roughly lines up with the collection of 36-month survey data for the individuals who were randomly assigned late in the intake period, but again may not contain earnings data for the full follow-up period for these individuals.

Despite the possibility of truncated earnings data for those participants who were randomly assigned late, we will use the data collected to the extent possible in the analysis of outcomes of interest along with the relevant survey data. In addition, the employment and earnings information from the 12 months before random assignment will be used as covariates in the analyses of impacts.

Accessing the state UI records requires several steps. First, the study team contacts states to request the data. Then a Memorandum of Understanding between the study team, DOL, and the state is signed, specifying the data request, confidentiality procedures, and compensation to the state for extracting the data for the study's use. Next, the study team sends the SSNs of study participants to the relevant state. State staff matches the provided SSNs to their respective wage records and compile a data file containing all the individuals they were able to match. This is then returned to the study team.

If DOL chooses to see access to these data via the NDNH, then we will assist ETA in preparing a formal data request to ACF, as Abt staff have done on two other recent evaluations conducted for DHHS. If successful, this request would produce data ready for analysis at a somewhat longer lag than with state-supplied data (since the NDNH has to be compiled nationally) but with the added benefit of a consolidated data set, reduced access costs, and coverage of federal employment.

4.3 Data Collection for the Process Study

Section 5 outlines the key research questions that will be addressed in the process study. In this section we describe the data we will collect for this aspect of the evaluation.

4.3.1 Interviews with Program Staff and Partners

Interviews with administrators and staff (including instructors and counselors) at each site will document the program services provided to the treatment group. These interviews will collect detailed information on a full range of process study topics including: program and organizational structure; service delivery systems; recruitment; nature and content of training and support services; key partners; linkages with employers; programmatic priorities; funding resources; sustainability of the grant program after the grant period; and the economic and programmatic context. Our overriding aim is to gain a complete understanding of the range of factors (programmatic, institutional, and economic) that serve to facilitate

⁵ If, in our preliminary discussions with states, we discover that they archive their files more frequently than every 45 months, we will either work out a process for accessing archived data or adjust the data collection schedule.

or inhibit the successful implementation and operation of the program. These interviews will also allow us to identify and obtain some information on other programs and services that may be available to study participants in the control group.

The research team will also conduct interviews with key program partners (e.g. one-stop centers, TANF agencies, community colleges) to help us understand the historical aspects of the partnership, the current relationships among different collaborating organizations, and the range of services provided. Finally interviews with two to three key employers from relevant sectors will be conducted to help us understand the extent to which critical “demand side” considerations have been integrated into the program model. The interviews will include discussions of employers’ roles in the planning process, their roles in program and curricula design, and their experiences with placement, hiring, and post-program employment of participants

Site visits will be a key method for acquiring data from interviews on the implementation and operations of the program. Researchers will use prepared discussion guides to conduct the semi-structured interviews, and will be guided by a set of protocols that cover the types of information required to advance our understanding of the training programs. The guide will be an outline of key topics of interest with sample questions and probes. The semi-structured nature of the interview guide will be purposively designed to allow site visitors maximum flexibility in tailoring their discussions during specific interviews to the different perspectives of respondents and the unique circumstances that prevail at each site while still ensuring that all key topic areas of interest are addressed in each site visit. While we will try to capture as much detail as possible, the questions in the instruments will remain open-ended in style to allow for the greatest freedom for the respondent to answer in his or her own words. The team will work closely with the sites to arrange for the most convenient but expeditious time to visit their program. We will also hold a site visitor training for all staff involved in conducting the visits. After each site visit, the data and information collected will be summarized and maintained in site specific databases.

Site visits will be conducted by two-person teams. Each team will be led by a senior researcher, joined by a mid-level/junior researcher, all of whom have experience in conducting site visits to educational and employment programs. To the extent possible, the teams will maintain site assignments over the course of the evaluation.

We will conduct two rounds of field research visits to each site. The first round of visits will be conducted approximately six months after the start of random assignment and will last three days. These visits will focus on documenting the implementation of the programs and will include interviews with administrators, staff, partners, and employers as well as focus groups. The second round of site visits will occur approximately 14 to 18 months after the start of random assignment when programs have reached maturity, and will focus on changes and developments in the provision of services as well as issues regarding the sustainability of the grant program. Given that we will already have a basic understanding of the program and its operation, these visits will be two days in length.

4.3.2 Program Documents

The grant application, policy and procedures manuals, staff training materials, recruitment materials, curricula, aggregate statistical reports, and other documents are an important source of information on program design and operating strategies and on the policy context (e.g., other programs supporting the working poor) for the process study. Published labor market information will also be used.

4.3.3 Program Enrollment, Attendance, and Completion Data

Administrative records from the grantee-funded program will be used to provide more detailed attendance and completion dates (beyond what can be collected through the survey) for the treatment group—and for the control group, to the extent that its members participate in non-grant-funded education and training at the same institutions and have that attendance logged in the same data system as grant-funded services. The precise data system involved will vary by site because we will rely primarily on data systems used by grantee staff, rather than imposing new forms or data systems specifically for the evaluation. When available from existing systems, key information that will be collected include: program and/or course enrollment, hours attended, program completion, credential receipt, credit receipt, receipt of support services (counseling, child care, transportation, financial assistance), and receipt of financial assistance.

The first step will be to assess availability and feasibility of collecting this information, which is likely to vary depending on the site. If there are a wide range of education and training providers in a particular community, this approach may provide very limited information on the types of programs in which individuals enroll beyond the grant program. However, in some cases, the institution where the grant-funded program operates may be the most likely place that both treatment and control group members would receive other types of education and training (beyond the grant-funded program), and accessing their administrative programs records may yield useful data on program enrollment and completion for both research groups. Additionally there may be only a limited number of places where the control group could seek out services, and thus this approach would provide useful data.

4.3.4 Participant Focus Groups

The participant perspective will be critical to understanding service utilization, reasons why services are or are not successful in achieving their goals, and insights on job advancement or job loss. Thus, to supplement the survey, we will conduct in-depth focus groups with small numbers of students in the treatment group as part of the first round of site visits.

Section 5 – Process Analysis

The process analysis for the Green Jobs-Health Care Impact Evaluation will serve four key purposes: (1) to describe the program design and operations in each site; (2) to help interpret the impact analysis results; (3) to understand the potential for pooling across sites; and, (4) to identify lessons learned for purposes of program replication. This section lays out the key research areas for the process analysis; the program dimensions to be examined; the data sources we will utilize to answer these questions; the timeline for data collection; and the methods we will use to analyze the data.

5.1 Research Areas

Corresponding to its four purposes, the process analysis will focus on the following overarching research areas:

- ***Description of program design and operations in each site.*** The process analysis will describe the program design. Because the program as it is described “on paper” (in the grant application or other materials) may differ from the actual program being tested, the process analysis will also describe the program as implemented. As detailed below, the process analysis will document the following for each site: program design and operations, local context, service receipt and utilization, sample members’ views and perspectives, and implementation accomplishments and challenges.
- ***Examination of treatment-control differential to help interpret impact results.*** Impacts on employment patterns, job characteristics, earnings and other outcomes will presumably be driven by *differences* in the amount and/or types of training and other services received by members of the treatment and control groups. Because the control group can access training and other services outside the grant-funded program, the process analysis will look to describe and establish levels of service *receipt for both treatment and control group members*. We will collect information on other sources of similar training (including those within the same institution) and sources of funding for training (e.g., other WIA programs).
- ***Identification of lessons learned for use in replication efforts.*** The process analysis will serve as a “road map” to aid policymakers and program administrators interested in developing similar approaches. Data from the process analysis—considered within the context of the impact results—will be the key source for formulating recommendations about how to replicate successful programs or improve upon their results. These data will also be used to identify lessons learned about the relative effectiveness of particular training strategies. While it may not be possible to completely disentangle which factors are driving differences in impacts across sites, to the extent possible, the analysis will identify factors that appear to be linked to success, as well as those that are not.

5.2 Key Program Dimensions

To inform the process analysis work, we will examine five key program dimensions: program design and operations; local context; service receipt and utilization; participant perspective; and implementation accomplishments and challenges. This section discusses each of these dimensions in turn, and Section 5.3 discusses the data sources that will be used to examine each dimension.

5.2.1 Program Design and Operations

A major contribution of the process study will be the description of program services—in particular, what the services actually look like and how they are delivered. The process study will collect detailed information on the length and sequence of training components and supports provided to students to encourage completion of training and job placement, retention, and advancement. While the specific nature of the issues addressed will depend on each program’s specific approach and services, the questions outlined in Figure 5-1 will be examined across all sites:

Figure 5-1. Program Design and Operations Dimensions

Program goals
- What are the goals of the program?
- To what extent does the program focus on educational attainment, employment, and employment retention and advancements?
- How and why were they established? Who established the goals?
- What were the site’s performance goals? To what extent did it reach these goals?
Target group and recruitment
- What is the target group?
- How do the demographics of the target group “match” to the program services?
- How are members of the target group actually identified and informed about available services?
- How are potential participants recruited?
- Are partnerships established to help recruit students?
Organizational structure and institutional partners
- What organization(s) are involved in planning, funding, and providing services?
- What is the role of the one-stop delivery system?
- What are the organizations’ linkages with one another? employers? and other service providers?
- What is the operational structure and size of organizations involved in the program?
- What is the length of time that the organization has been providing services and how established is the organization within the community?
- How is the program staffed?
Curriculum and training strategy
- What is the nature (e.g. work-based, classroom) and content of the training provided in terms of occupational focus, courses, and time required to complete training? What credentials are connected to the training? Are the credentials recognized by employers?
- What are the career pathway components of the program (links to next steps on the training ladder, connections to employers, open entry/open exit)?
- Are the courses for credit?
- How are the curricula contextualized (i.e. integration of basic skills and training)?
- Are there direct connections to employment, and links to other steps on a career path?
- Are courses “chunked” or sequenced to accommodate working adults?
- To what extent was the program (or any components) newly established through the grant program or operational prior to the grant award?
Connections to employers and employment
- To what extent were employers involved in developing the training program?
- Did they play a role in course development or curricula design?
- Are they involved in hiring individuals after training is completed?

Counseling and other support services

- What additional supports are available to students, such as assessment, case management, career counseling, child care, transportation assistance, financial assistance, and supports with personal issues?
- What type and how many staff are available to provide these services?
- How are participants chosen to receive additional supports?

5.2.2 Local Context

The local context will shape the opportunities and incentives facing program participants and, hence, the implementation and impacts of the programs. The issues examined in this area include: (1) the labor market context, particularly for the industry in which the training is being provided, such as job availability and characteristics of available jobs); (2) education and economic development policies including the Workforce Investment Act and post-secondary education (e.g., community colleges); and, (3) the range of available supports for low-income individuals including Pell Grants and other financial assistance, child care, SNAP, and barrier-related services.

5.2.3 Service Receipt and Utilization

Primarily using the 18-month follow-up survey, the study will tabulate the amount and type of services received by members of each research group. This quantitative analysis will help the evaluation to understand the nature of the treatment difference as described in section 5.1. For both research groups, both the 18-month and the 36-month follow-up surveys will measure whether the individuals receive training; what type of training is provided; what “dose” of services they receive, particularly rates of utilization of various services and supports, and completion rates. In addition, we will explore whether the National Student Clearinghouse would be a useful source of data on enrollment and degree completion for both treatment and control group members.

For the treatment group, if feasible and depending on the accessibility of grantee-level program data, more detailed data on participation patterns will be collected on training and services received via the grantee’s administrative data records. It will also be important to document when the grant program ended, and determine if and how this affected the receipt of training services by the treatment group.

5.2.4 Participant Views of Services and Barriers

Documenting the participant perspective is critical to understanding service utilization and the reasons why services are or are not successful in achieving their goals. We will explore what treatment group members hear and know about programs and services; reasons individuals use or do not use program services; particular challenges they may face in attending or completing school; participant knowledge of available resources; and perceptions about the likelihood of advancement.

5.2.5 Implementation Accomplishments and Challenges

The implementation study will document the accomplishments and challenges of the participating programs, with the specific goal of assisting other program administrators in designing and replicating the training program models without reinventing the wheel. Key questions to be addressed include: What were the primary successes and challenges, and how did they impact the delivery of services and achievement of program goals? What contributed to these successes and challenges? Was the program able to effectively engage employers from the relevant industry? How? How did the programs address the key challenges that have hindered past programs? What strategies do they use to engage families and facilitate their attendance and completion (e.g., linkages to employers, contextualized curricula, personal and vocational supports, location and hours of classes, nature of recruitment “pitch”)? Topics addressed

will include policy, budgetary, institutional, and organizational issues as well as the contextual issues discussed in 5.2.2.

5.3 Use of Data Sources

As described in Section 4, we will use a wide range of data sources to explore these research areas and program dimensions. Figure 5-2 shows the data sources for each program dimension.

Figure 5-2. Process Study Program Dimensions and Data Sources

Program Dimension	Data Sources
<p><i>Local context:</i> Broad community context in which the program operates/services are delivered</p> <ul style="list-style-type: none"> Socioeconomic and ethnic profile of the population Unemployment rates, availability of jobs, characteristics of available jobs Range of education and training opportunities in the community Availability of public and financial supports 	<p>Documents</p> <ul style="list-style-type: none"> Census data Public information and advocacy materials BLS area employment and earnings data Program plans and manuals <p>Informants</p> <ul style="list-style-type: none"> Program staff Program partners
<p><i>Program design/operations:</i> Characteristics of the grantees: organizational characteristics, staffing, program partners.</p> <ul style="list-style-type: none"> <u>Organizational Structure:</u> size, operational structure, funding, history, leadership, linkages with other systems (local work force boards, community colleges) <u>Staffing:</u> Number and roles of staff (planned and actual) <p><u>Partners:</u> Program services offered and delivered, how services are coordinated</p>	<p>Documents/Data</p> <ul style="list-style-type: none"> Grant application Program plans and manuals Training records <p>Informants</p> <ul style="list-style-type: none"> Program administrative staff Program service delivery staff (teachers, employers, counselors, other professionals) Program partners and employers
<p><i>Program design/operations:</i> Strategies used by the program to deliver curricula/services or organize activities</p> <ul style="list-style-type: none"> Outreach and recruitment strategies (planned and actual) Assessment and case management Course requirements and schedules Instructional methods and curricula Counseling and other support services Location of services, activities Role of employers 	<p>Documents</p> <ul style="list-style-type: none"> Newsletters, recruiting materials, brochures, program planning documents, print and electronic products Course curricula <p>Informants</p> <ul style="list-style-type: none"> Program staff, program partners
<p><i>Service receipt/utilization:</i> Services received by treatment and control group members</p> <ul style="list-style-type: none"> Number and type of services received Length of participation in services Completion and credential receipt Other education, job training, and support service programs available 	<p>Documents</p> <ul style="list-style-type: none"> Program plans and manuals <p>Informants</p> <ul style="list-style-type: none"> Program staff Program enrollment and completion data (treatment group only) Follow-up survey Participant focus groups
<p><i>Participant perspective:</i> Factors that affect use/non-use of services</p> <ul style="list-style-type: none"> How heard about services/messaging Challenges/facilitators to using services 	<p>Informants</p> <ul style="list-style-type: none"> Participant focus groups Follow-up survey
<p><i>Implementation accomplishments/challenges:</i> Factors that facilitated or impeded the effective delivery to services to participants</p>	<p>Documents</p> <ul style="list-style-type: none"> Needs assessments Planning documents <p>Informants</p> <ul style="list-style-type: none"> Program staff, partners, employers

5.4 Analysis Methods

Information collected for the process study will be integrated and used to develop a database on program operations for each site. Data on the key program dimensions will be used to describe each program model and the treatment/control group differential, and help interpret the impacts that are observed. The process study will use multiple analytic approaches. To synthesize a coherent “story” about program operations, for example, we will use *narrative description* of data from interviews, focus groups, and documentary materials. To assist with these analyses, we will load all information collected into NVivo (software developed by QSR International). NVivo tools can be used to organize and analyze qualitative and unstructured information material. In contrast, to measure quantifiable aspects of the treatment—from the surveys, and structured observations—we will use *descriptive statistics*.

The analysis of qualitative data across sites will be a critical component of the process analysis. In a multisite case study, it can be particularly challenging to compare programs that exist in very different contexts and adopt different approaches. We will use the following methods to organize and analyze the qualitative data: (1) create a logic model, or “theory of change,” based on a conceptual framework that is site-specific, detailing the processes through which changes in outcomes are theoretically expected to occur given the nature of the intervention and target population in a given site; (2) create detailed descriptions of each site, including an analysis of the issues or themes that each site presents, and that the sites as a group present; (3) focus on describing specific program dimensions identified above, including analytical tables that systematically organize the key preliminary findings for each site under each main research area and quantify aspects of program implementation; (4) detail site timelines or chronologies of the key stages of program development and implementation; and, (5) map participant flow through the program.

The analysis phase should include sorting the data in many different ways to expose or create new insights and looking for conflicting data to disconfirm the analysis. Several analytical techniques will be used to identify findings and work toward conclusions. These strategies help researchers to move beyond initial impressions to develop accurate and reliable findings. These strategies include: (1) comparing multiple sources of evidence before deciding there is a finding; (2) analyzing information within each site for themes and then across all sites for themes that are either the same or different, (3) examining how data collection and analysis findings compare to original expectations and hypotheses, (4) analyzing all the information collected to develop a picture of what is happening and why.

Based on these analyses and on the similarities and differences across programs in their design, implementation, and impacts, the team will explore whether there is any evidence suggesting that certain training models or program features lead to larger impacts, for potential testing by later research projects capable of establishing causal connections (e.g., by varying these factors randomly across sites or participants). Our ability to do this will be highly limited, however, by the inclusion of only four training programs in the evaluation. The team will also describe the main successes and challenges identified by site respondents and the lessons learned. Finally, we will seek to identify potentially promising practices that are important for those administrators interested in replication. These analyses will be used as the foundation for implementation reports which will provide useful insights to a broad audience of researchers, policymakers, and practitioners.

Section 6 – Impact Analysis

The primary purpose of the impact analysis for the Green Jobs and Health Care Evaluation is to estimate program impacts—that is, outcomes for the treatment group relative to what their outcomes would have been in the absence of the program—in each of the four programs studied. This section lays out the research questions for the impact analysis, describes our analytic approach to answer those questions, and examines the statistical precision of the answers we will obtain to determine the smallest true impacts that can be confidently detected given the study design (i.e., the minimum detectable impacts).

6.1 Impact Research Questions

Of central importance to the success of each studied program is the effect of access to training on participants' economic outcomes. Due to both public policy considerations and analytical considerations that we discuss below, it is valuable to define a single outcome of primary interest (a single “confirmatory outcome”), while still reporting other outcomes of secondary importance in the evaluation (“exploratory outcomes”). Towards that strategy, we plan to focus the evaluation's impact analysis on a single confirmatory research question: *What is the impact of access to grantee services on earnings?* We will supplement this research question with other ones related to other outcomes of interest, including participation in training, attainment of credentials, and employment.

In this section we discuss several issues related to the research questions for the impact analysis. Section 6.1.1 explains why we use a single confirmatory outcome. Section 6.1.2 justifies our particular choice for that confirmatory outcome. Section 6.1.3 explains the tradeoffs associated with measuring that outcome with survey or with administrative data. Section 6.1.4 discusses the other, “exploratory” outcomes used for the impact analysis. Section 6.1.5 explains the difference between estimates of the effect of *access* to grantee-funded services—conventionally known as intention to treat (ITT) estimates—and estimates of the effect of *actually participating* in those services when some individuals granted access do not participate—conventionally known as the effect of treatment on the treated (TOT) estimates.

6.1.1 Primary Confirmatory Outcome

When seeking to determine the overall effectiveness of an intervention—i.e., asking whether a program has a favorable impact as opposed to no impact—testing for impacts on a variety of outcomes or for a variety of participant subgroups is statistically problematic. Even if there are no true impacts, the likelihood of finding at least one statistically significant effect and therefore rejecting the null hypothesis of no impact increases rapidly with the number of tests, to well above the intended 5 or 10 percent level. This situation, referred to as the “multiple comparisons” problem, can arise either when different research questions are asked for a single site (that is, different outcomes that are examined for the same study sample) or when a single question is examined across different sites or for different subgroups (that is, a single outcome with different study samples). Statistical techniques can be used to adjust results to take into account the higher likelihood of detecting a spuriously significant result when multiple comparisons are involved.

Following Schochet (2009), we will address the multiple comparisons problem in this evaluation by identifying a single outcome as the primary confirmatory impact analysis: post training earnings in the third year after random assignment. To understand this choice, we note that in general, post-training earnings is the critical outcome of interest, not only to participants in the grant-funded training but also to DOL and to society at-large since it represents the principal policy goal of the ARRA grant programs—to increase participants' earnings on a sustained basis. The specific measure of sustained earnings could

take a variety of forms. Discussion with DOL leads us to conclude that total earnings in months 25-36 after random assignment is a strong candidate for the confirmatory outcome, even though a broader measure of earnings aggregated over months 1-36 after random assignment would also be a reasonable choice.

Training is expected to last four to six months for many treatment group members, though it will be shorter for some treatment group members and longer for others (e.g., those participating in long-term training at two-year community colleges). Moreover, it seems likely that, because time spent in training will reduce availability for work, the treatment group's earnings during, and possibly shortly after, training could be lower than is the case for the control group.⁶ While we will report impacts over the entire follow-up period, beginning at random assignment in the exploratory analyses described below, to test the key research question of whether sustained, post-program impacts are attained as a confirmatory analysis we will exclude from the confirmatory outcome measure earnings during and shortly after when participants are likely to be in training. This suggests waiting as far into the follow-up period as possible to measure earnings for the confirmatory analysis while avoiding time intervals so short that seasonal variation in earnings come into play. For this reason, earnings in the latest possible 12-month period, months 25 through 36 after random assignment, seems the best indicator of long-run earnings gains.⁷ In addition to gauging earnings impacts, the evaluation will consider impacts other outcomes—such as educational attainment, credentialing, and job quality— as part of an exploratory analysis described below because of the role of these factors in influencing earnings outcomes.

We will use a two-tailed test, setting our threshold for statistical significance at 10% for the confirmatory outcome. In our view the main alternative to 10%--a 5% significance level—when combined with the multiple comparisons adjustment discussed below, would tilt the confirmatory hypothesis testing too much in the direction of avoiding false positive results (Type I error) while unduly expanding the risk of false negatives (i.e., a finding of no significant impact when an effect of important magnitude in fact occurs—Type II error). Because prior research has shown some negative effects of training (at least over some time periods and for some subgroups), we think it prudent not to impose a directional hypothesis and use a one-tailed test. We expect also to report the confidence intervals to convey the degree of certainty and precision with which the results should be interpreted.

Although focusing on a single confirmatory outcome avoids potential problems that would arise from use of more than one outcome of primary interest, it does not completely avoid the multiple comparisons problem because the evaluation plans to estimate site-specific impact results for each of the grantees in the study. As described in more detail in Section 6.2, the evaluation will estimate four separate confirmatory impacts—one for each of the four study sites. The goal of this analysis will be to determine if positive earnings impacts have occurred in one or more sites and, if so, to identify the specific site-level intervention or interventions that produced effects. We expect to make multiple comparisons adjustments for the confirmatory impact results because, with four study sites, the likelihood of finding a

⁶ This finding has appeared in the literature on the impacts of training in other contexts involving disadvantaged workers. See for example Greenberg, Michalopoulos and Robins (2004) or Hotz, Imbens and Klerman (2006).

⁷ We do not want to extend the confirmatory analysis into even later months through extrapolation beyond the observable data because of the uncertainty that would arise with the extrapolation process (Greenberg, Michalopoulos and Robins, 2004).

statistically significant impact for one or more of them by chance alone, when no true impacts have occurred in any site, is far greater than 10 percent.⁸

A large literature exists regarding the best way to protect against false positives of this sort when doing multiple tests of potential impacts, summarized in Schochet (2009). A straightforward, though conservative method, is the Bonferroni adjustment which uses a much smaller significance level (i.e., an alpha well below .10) in conducting each individual confirmatory test. In particular, this method would use a significance level of .025 for each of the four confirmatory tests, thus ensuring that at most the chance of spuriously significant results for one or more of the tests is .10 (.025 x 4). The method is conservative for two reasons: it assumes that the cumulative risk is the sum of the individual risks, which overstates the case given that once one false positive has occurred subsequent false positives do nothing to heighten the overall risk of one or more such occurrences. Moreover, it fails to recognize that ordering four impact estimates from largest to smallest lowers the risk of a spuriously significant result occurring on the second test given that none occurred on the first test—and similarly for the third and fourth tests. A procedure developed by Benjamini and Hochberg (1995) takes this factor into account and thereby lowers the probability of Type II error—i.e., failure to detect non-zero impacts that do occur—while restraining the false positive rate to be .10 or less. For power analysis purposes (i.e., examination of minimum detectable impacts) later in the chapter, we assume that the Bonferroni adjustment will be applied to the four confirmatory earnings impact tests since any procedure that is used will achieve at least that degree of statistical power. The adjustment procedure we actually apply will depend on developments in the literature as statisticians continue to find ways to increase statistical power when testing multiple hypotheses. Our current plan is to use the Benjamini-Hochberg procedure, but we will continue to monitor the literature for further developments in the field. If a more powerful procedure achieves widespread use, we will revise our plans and adopt it.

The previous discussion concerns the Final Report. Because we have defined the study's confirmatory outcome as long-term earnings measured in the third year post-random assignment, we will not be able to report on the "success" of the intervention in the Interim Report, which will examine earnings only in months 1-18 following random assignment. Instead that report will focus on implementation and participation and cover early exploratory outcomes, with emphasis on the point that judging the intervention's overall success must wait until month 25-36 earnings impact estimates become available in the Final Report. In particular, in the executive summary and conclusions chapter of the Interim Report—and near the beginning of each chapter of impact findings—we will include a statement along the lines of "None of the findings in the current Report provide *proof* of the effectiveness of the training programs studied, not even the statistically significant impact findings presented. By prior specification of the research protocol for the evaluation,⁹ the only finding that can provide proof of effectiveness is the impact of the training programs on the earnings of treatment group members in months 25-36 after random assignment, information not yet available but that will be presented as part of the evaluation's 'confirmatory analysis' in the study's Final Report. All other impact findings from the evaluation,

⁸ Indeed, that probability equals 1 minus .9 to the fourth power, or .34 (i.e., 34 percent).

⁹ A footnote will appear in the Interim Report text at this point to the effect that "See *Green Jobs and Health Care Impact Evaluation: Evaluation Design Report* (Abt Associates, 2011) for an explanation of this protocol, including the statistical reasons that analyses seeking to prove that one or more of the four training programs studied improved participants' lives are to be based on a single outcome measure in order to minimize the chance of drawing spuriously favorable conclusions about the interventions when many different impact estimates are tested"

including the statistically significant findings in the current Report, come from ‘exploratory’ analyses that can *suggest*—but not prove—that beneficial effects of the training have occurred in other realms, including earnings in months prior to month 25 following random assignment.” Similar language will be provided when exploratory impact findings are presented in the Final Report

6.1.2 Data Sources for the Confirmatory Outcome Measure

A remaining issue to consider is which source of follow-up data to use for our primary confirmatory outcome measure, given that we plan to collect both administrative and survey data on the employment and earnings of study participants. Here we discuss the major advantages of each data source relative to the other, as summarized in Figure 6-1 and addressed to some extent in the employment and training literature (see for example Kornfeld and Bloom, 1999).

Figure 6-1. Advantages of Administrative and Survey Data on Annual Earnings

Area of Relative Strength	Administrative Data	Survey Data
Completeness of Analysis Sample	100 percent complete rather than 80 percent → smaller MDIs → no threat of sample omissions (i.e., survey non-response bias)	--
Measurement Accuracy	No over- or under-statement of earnings due to recall error Consistent measurement of earnings on a pre-tax basis inclusive of tips and bonuses	No “false 0s” due to <ul style="list-style-type: none"> • omission of federal employment* • omission of self-employment • omission of informal employment • out-of-state employment* • incorrect Social Security numbers Greater consistency with related exploratory outcomes (e.g., weeks worked) Aligns exactly with months 25-36 after random assignment
Data Availability	--	Shorter lag from end of follow-up period to data availability
*Not an advantage over administrative data on quarterly earnings obtained from the NDNH rather than state Unemployment Insurance agencies (see Chapter 4 for a discussion of these two earnings data sources and the differences between them).		

As shown at the top of the figure, one advantage of administrative records is that they produce analysis data for 100 percent of the sample, compared to survey data which we expect to cover about 80 percent of the sample (assuming 20 percent of the combined treatment and control groups proves untraceable or unwilling to be interviewed). We are not saying that *records of positive earnings* will be obtained for all sample members from administrative sources. Rather, we—in line with previous earnings analyses based on administrative data—will assume that the absence of a record of earnings for a given person in a given calendar quarter means that that individual earned \$0 during that period, rather than that s/he had missing data on earnings. (We discuss the disadvantages of this assumption shortly.) This will put all individuals who are randomly assigned into the analysis sample.

Given this framework, there are two consequences of the more complete sample coverage provided by administrative data. First, minimum detectable impacts (MDI) for the annual earnings confirmatory

outcome will be smaller if we use an administrative data sample that exceeds in size that of the follow-up surveys by a ratio of 5-to-4. It will not be dramatically smaller, however; about 11 percent smaller MDIs result from shifting from 80 percent non-missing data to 100-percent non-missing data (see section 6.2 on MDIs below). An additional advantage of the 100-percent sample coverage of the administrative data is avoidance of any risk of bias in the earnings impact findings due to non-coverage. The 80 percent of sample members who complete follow-up interviews might not be representative of all study participants the way administrative data are given the assumption of 100 percent coverage. Even so, before relying on survey data we can explore the extent to which non-response bias may be occurring in the survey data and use baseline and administrative data on all sample members to construct appropriate analysis weights for survey respondents in order to remove some of the survey non-response bias. Moreover, in as much as non-response is symmetric between the treatment and control groups impact estimates from the survey will be unbiased *for the respondent subpopulation*—but not necessarily for all sample members including survey non-respondents.

Administrative data also have two advantages over survey data in the area of measurement accuracy: they are not subject to recall error the way follow-up survey outcome measures can be, and they measure earnings on a consistent pre-tax basis inclusive of tips and bonuses. In administrative data there are no forgotten jobs, no misaligned time periods when reporting jobs and earnings, and no under- or over-statement of the number of hours worked or the pay earned for those jobs. Moreover, the proclivity of some individuals to self-report earnings on surveys as their post-tax take-home pay, and to omit tips and/or bonuses, is avoided with administrative data.

All the other advantages in Figure 6-1 lie with survey data column. In terms of measurement accuracy, a major advantage of survey reporting is that there are no “false 0s” in the earnings information unless jobs are forgotten or put in the wrong time periods. In particular, unlike administrative earnings data from state Unemployment Insurance (UI) wage reporting systems, we do not end up with quarters showing \$0 in earnings because federal employment, self-employment, informal employment, and out-of-state employment¹⁰ are omitted from state UI records, for which we infer \$0 earnings in quarters with no records showing up on our data extracts. Omitted jobs will artificially lower earnings to \$0 when these are the only jobs in a quarter—and lower the total earnings amount (which will still be greater than \$0) in quarters with multiple jobs. These problems constitute a smaller advantage for survey data compared to administrative data if quarterly earnings information is obtained from the national NDNH data system, which covers federal employment and out-of-state employment. The extreme case of omitted jobs and “false 0s” in the administrative data arises from incorrect Social Security numbers (SSNs) for sample members, since SSNs provide the principle link to earnings records for agencies seeking to fulfill data requests from evaluators

A further measurement advantage of survey data is the consistency between confirmatory earnings outcomes and related exploratory outcomes such as weeks worked and hours worked. The latter outcomes can only be measured on follow-up surveys, and are collected in a way that assures synchronization with total earnings measures for the quarter. Not so when quarterly earnings are taken from administrative data, give that source’s alternative sample coverage and job-type omissions. Finally, survey data have the measurement advantage of obtaining earnings information by job spell—which allows for calculation of *monthly* (or even weekly) earnings. Thus, if surveys are used to measure

¹⁰ Although none of the selected sites is located in a major bi- or tri-state area, the absence of information about out-of-state employment might be pertinent if study participants move away from the area in which grantee services are provided.

earnings we can use total earnings for precisely months 25-36 for each sample members as our confirmatory outcome. With administrative data, we could not as precisely measure this interval for the great majority of individuals not randomly assigned on the first day of a calendar quarter.

The final advantage of survey data concerns data availability. Survey earnings measures are available more quickly after the end of the follow-up period to be studied than administrative data. Complete administrative earnings data are not typically available for six months, or possibly longer, after the end of the last calendar quarter to be included in the impact analysis.

Other research has examined the differences in impact estimates arising from use of the two different data sources. For example, the national study of the Job Corps program found that the patterns in estimated impacts using the survey and administrative data were similar, but the survey-based estimates were generally larger and more likely to be statistically significant (Schochet, McConnell, and Burghardt, 2003). Although the potential reasons for the differences were numerous, two reasons appeared to dominate. First survey-based levels of earnings were larger, due in part to incorrect reporting of Social Security numbers by employers or sample members, noncoverage of some formal and informal jobs in the administrative data, and likely over-reporting of the hours worked in the survey data. Second, earnings impacts measured with the administrative data were larger for survey respondents than were impacts for non-respondents—which suggests that survey-based impact estimates were slightly biased upward. We cannot anticipate whether the same will be true in this evaluation, so for now the question of whether to use survey or administrative data for this study’s confirmatory outcome—earnings in months 25-36 of follow-up—is still open. We will make a decision on this matter once we have had a chance to examine short-term earnings information from both administrative records and the early follow-up survey ahead of conducting the Interim Report’s impact analyses. This will allow us to explore empirically how the different factors in Figure 6-1 are playing out in our particular sample with the particular earnings measurement methods employed by our survey questionnaires and thus to make a more informed determination of which source has the strongest advantages overall. To avoid “clouding” this decision with knowledge of how *measured impacts* differ between sources, the treatment/control status of sample members will be masked in the data sets used for this analysis. The Interim Report will declare our decision on which source will be used in the study’s only confirmatory impact analysis in the Final Report.

6.1.3 Exploratory Analyses

As explained earlier, while the primary confirmatory analysis will examine the program effect on earnings, the evaluation will also include exploratory analyses of a wide range of other outcomes and subgroups of interest to DOL. The exploratory analyses will augment the understanding of the main impact estimate by providing insights about how and why the main impact findings are what they are and whether the grant services were helpful for all or only a subset of study participants. Statistically significant exploratory findings will never be depicted as *proving* that training had an impact of was in general was a “success”; only confirmatory analyses that rigorously guard against “false positives” can do this. Instead, we will interpret statistically significant exploratory impact findings as *suggestive* of potential impacts that, based on the best available evidence, one could *hypothesize* have occurred.

Three types of outcomes will be examined in the exploratory analysis to obtain suggestive indications of where the training intervention may have improved the lives of treatment group members. First, we will examine the impact of greater service access on actual services received (from *all* sources, not just ARRA-funded grant services), including:

- participation in training,

- completion of training,
- participation in reemployment services, and
- receipt of a credential related to training.

Second, in addition to the primary confirmatory outcome of earnings in the third year after random assignment, we will explore earnings over various other time periods which have been considered by other evaluations of employment and training programs. These include total earnings over the full three year follow-up period and the time path of earnings by quarter during that period.

Finally, we will explore impacts on other employment-related outcomes such as:

- employment at any point in the follow-up period,
- employment status at the end of the observation period,
- employment in jobs with certain features, such as being in the occupation to which training pertained or one that offers fringe benefits or career progression, and
- participants' post-training economic security and financial stability, as indicated by household income, poverty status, financial hardship, and receipt of government-funded benefits.

We also plan to explore whether the impacts of access to grantee services varies by type of study participant.

Subgroups to be examined will be based on study participants' characteristics prior to random assignment, so that there is no confounding influence of the research group assignment on the characteristic. Because they are exploratory, these analyses can range across a variety of subgroup categories and outcome measure that are not specified in advance. The potential for spuriously significant effects when many subgroup outcomes are examined will be controlled not by statistical means and limited, pre-specified tests but by discipline in how we present and interpret results, as described for exploratory analyses in general above. In particular, we will explicitly note the multiple comparisons problem before discussing the subgroup results.

Examples of subgroups that may be part of the exploratory analyses include those defined based on the following traits of the study participants:

- demographic characteristics, such as their sex, race or ethnicity, and age¹¹
- prior education and employment, such as highest education level, employment status, earnings, and occupation
- receipt of government benefits, such as Temporary Assistance to Needy Families (TANF) and unemployment insurance benefits
- opinions about work, such as barriers to work, willingness to take different types of jobs, and self-efficacy

¹¹ We plan to examine the effects for those over age 21 (or over age 25), if sample size permits, in line with prior research that reveals variation in effects for youth and adults.

The estimation of impacts for subgroups will be part of the exploratory analyses. Because we will not claim to have proved an impact or difference in impacts occurred for any subgroup or subgroups, there is no need to adjust for the fact that multiple tests are conducted. We will have very weak power to detect impacts in subgroups and differential effects across subgroups in any case, given the small site-by-subgroup sample sizes anticipated for the evaluation. Thus, it will be important to not over-interpret the lack of significant variation in estimated impacts across subgroups as evidence that those impacts are highly uniform, and not to over-interpret the lack of significantly positive effects for individual subgroups as evidence that the studied training does not benefit them at all.

6.1.4 Principal Focus on the Effects of Access to Training

Both the primary confirmatory research question and the exploratory research questions focus on differences between treatment and control group members in their access to— rather than their participation in—grant-funded training. This is because the crucial difference between the two research groups is their *access* to grant-funded training and supportive services: individuals in the treatment group will have access to both grant-funded training and supportive services and other, potentially-similar services available in the community, while control group members will have access to only those other services in the community. In the evaluation literature, the impact estimate derived from this approach is referred to as the intent-to-treat (ITT) parameter. It measures the impact of offering the grant-funded service on the outcomes under consideration.

However, some treatment group members will not participate in grant-funded services despite their open access. And a small number of control group members may manage to participate in grant-funded services “through the backdoor” due to grantee error. In such circumstances, an alternative to the ITT impact estimate is an estimate of the impact of the treatment (grant-funded services) on the individuals who receive grant-funded services compared to what outcomes for those individuals would have been absent participation in those services—an approach that is known in the evaluation literature as measuring the treatment-on-the-treated (TOT) effect of an intervention. Our analysis will focus primarily on the ITT estimate rather than the TOT estimate because the ITT estimate is the more policy relevant concept: since the policy choice is whether to offer someone the training or other services, it reflects a world where such services remain voluntary; it would be neither possible nor desirable to make participation in training mandatory. Thus, knowing that completing the training has an impact on participants’ outcomes does not help with the policy decision of whether or not to offer training. (In addition, grant programs similar to those under evaluation presumably would be offered in addition to existing programs, so it is not policy-relevant to compare them to no training, even if we could.)

Nevertheless, we will also report some TOT estimates. We discuss how we will recover TOT estimates from the ITT estimates in section 6.3 on estimation methods below.

6.2 Minimum Detectable Impacts (MDIs)

MDIs are the smallest true impacts that the study has a high probability of detecting; the greater the statistical power of the design, the smaller the MDI. It is important to calculate MDIs before beginning an evaluation in order to ensure that the study will be able to detect impacts of magnitudes that are relevant to policymakers. In this section, we first present MDIs for three representative outcomes of interest, given our projections of likely sample sizes at each grantee.

MDIs are a function of several factors including the ratio of treatment to control participants, the standard deviation of the outcome being examined in the absence of the intervention, and, crucially, the sample size on which the analysis is conducted. For the GJ-HC evaluation, the relevant sample sizes are the

separate sample sizes for each grantee, rather than summing across grantees. This is because the four sites for the evaluation were selected purposively and not at random; therefore, the analysis will not be representative of grant recipients. We will analyze and report the findings from four independent tests of specific interventions, and in our confirmatory analysis make a multiple comparisons adjustment to the four tests of impacts on month 25-36 earnings. Our MDI calculations take this latter point into account.

The first two columns of Figure 6-2 present the projected treatment and control group sample sizes for each of the four study sites. The projected sample sizes are based on the study team's negotiations with the four grantees. These projections assume a 12- to 18-month random assignment sample intake period as needed.

The fourth column presents the MDIs for the primary confirmatory outcome, total earnings in months 25-36-after random assignment. The final two columns of the figure present MDIs for two illustrative exploratory outcomes of interest—the likelihood of employment and possession of a degree or credential in the 36th month after random assignment.

All the MDI calculations in the figure are based on a number of assumptions, some of which vary by site and by the outcome measure involved. These assumptions are as follows:

- the treatment-control ratio (which, as noted in earlier chapters, varies across sites) is maintained as constant throughout the sample intake period in any given site;
- the target sample sizes given in earlier chapters are achieved in each site;
- the standard deviation of annual earnings for males is \$16,000 and for females is \$11,000. (Average annual earnings are expected to be \$14,000 for males and \$10,000 for females.)¹²
- the standard deviation of annual earnings for the entire sample in any site varies across sites because of different anticipated gender compositions in different sites; based on information from site staff, the percentages of the sample who are male are assumed to be 20% for AIOIC, 40% for Grand Rapids, 20% for North Central Texas, and 95% for Kern;
- the share of the control group employed in the 36th month after random assignment is 65 percent;¹³
- the share of the control group with an educational degree or training credential in the 36th month after random assignment is 30 percent;¹⁴

¹² These assumptions are based on results from similar studies of similar interventions such as the Sectoral Employment Impact Study (Maguire et al., 2010), the National JTPA Study (Bloom et al., 1993), and the Welfare-to-Work Voucher evaluation (Mills et al., 2006).

¹³ This employment rate is based on the results of the National JTPA Study.

¹⁴ This rate of degree or credential attainment comes from the baseline sample in the ITA Experiment, where 25 percent of individuals who wanted to receive training had a degree or credential. Since this was a baseline measure, a rate of .30 by the end of the follow-up period was assumed for these MDI calculations. (Note that the ITA experiment did not have a no-treatment control group, which is why baseline rates are used.)

- the inclusion of baseline characteristics of sample members as covariates in the impact regressions will account for approximately 20 percent of the total variance in outcomes¹⁵;
- the follow-up surveys used to measure outcomes for the impact analysis will achieve an 80 percent response rate;¹⁶ and

For the confirmatory outcome, annual earnings in months 25-36, a Bonferoni adjustment is made to the threshold of statistical significance in the MDI calculations, lowering the p-value threshold for rejecting the null hypothesis of no impact on earnings in a given site from 0.10 to 0.025 to ensure the overall probability of a “false positive” statistically significant impact finding from the four sites combined does not exceed 0.10. This is a conservative approach since, as noted above, the specific procedure used to adjust for multiple confirmatory tests will depend on the best—i.e., statistically most powerful—methodology available in the literature when the first impact analysis is conducted; we therefore expect the true MDIs for annual earnings to be smaller than those shown here.

Figure 6-2. MDIs for Confirmatory and Exploratory Outcomes of Interest

	Treatment sample size	Control sample size	MDI: annual earnings	MDI: Employment rate	MDI: Degree or credential
AIOIC (MN)	600	600	\$2,099	7.7%	7.4%
Grand Rapids (MI)	600	300	\$2,801	9.4%	9.1%
North Central Texas	589	485	\$2,229	8.2%	7.9%
Kern (CA)	425	425	\$3,215	9.2%	8.8%

An important factor in the MDI calculations is the target sample sizes for the treatment and control groups. At the site with the largest expected sample sizes, AIOIC, our analysis suggests that we could detect an impact of \$2,099 in annual earnings, a 7.7 percentage point impact on the employment rate, and a 7.4 percentage point impact on holding a degree or credential. Sample size projections for the smallest site, Kern, are shown in the final row of the figure, with 425 participants in the treatment group and 425 in the control group. While different gender mixes of anticipated participants also cause MDIs for earnings to differ across sites, sample size differences cause most of the variation seen in the figure. Hence, MDIs for Kern are larger than those at AIOIC primarily as a result of smaller samples, increasing to \$3,215 in annual earnings, 9.2 percentage points of employment, and 8.8 percentage points for degree or credential attainment. Generally speaking, the MDIs for North Central Texas are fairly close to those

¹⁵ Previous studies of impacts on the earnings of disadvantaged groups using similar baseline characteristics have explained around 20 percent of earnings variance from those characteristics. For example, the recently published Sectoral Employment Impact Study (SEIS), a random assignment study of an intervention for under skilled, unemployed, and low-income adults, reported explanatory power of 14 to 19 percent, while the National Job Corps Study achieved 20 percent on this measure.

¹⁶ Although we might choose to use administrative data with 100 percent coverage for the annual earnings and employment rate outcomes, to be conservative we present MDIs here assuming the use of survey data. If administrative data are used instead, MDIs for annual earnings and employment rate will be 11 percent smaller than shown here.

for AIOIC, given that their expected sample sizes are fairly close; the MDIs for Grand Rapids are more comparable to those of Kern.¹⁷

These MDIs are roughly in line with the magnitude of impacts found in two recent studies of employment and training programs, though at the high end of the range of impact magnitudes found in earlier randomized impact studies of such programs. For example, the recently published Sectoral Employment Impact Study (SEIS), a random assignment study of an intervention for underskilled, unemployed, and low-income adults, found impacts on earnings in the second year of follow-up of \$3,777 for men and \$4,555 for women. Another study, the Workforce Investment Act Non-Experimental Net Impact Evaluation (Heinrich, Mueser, and Troske, 2009), which examined the effects of WIA-funded services on dislocated workers and adults who were generally low-income, found a difference of approximately \$450 in quarterly earnings for men who participated in WIA training and \$650 for women at six quarters after program entry; these roughly correspond to annual earnings differences of about \$1,800 for men and \$2,600 for women.¹⁸

Although neither of these studies is exactly the same as the GJ-HC Evaluation in terms of the program model, the community and labor market contexts, and the target populations, they are similar enough to be broadly comparable to the current study. For instance, SEIS examined three programs that offered a combination of short-term training and job placement assistance for unemployed and low-income adults. The WIA Non-experimental net impact evaluation examined the broad population of adult and dislocated workers seeking WIA Title I training services, which are typically short-term trainings working toward an occupational credential.

However, it is important to note that one very prominent evaluation of training programs for disadvantaged adults in the late 1980s, the National JTPA Study, found impacts of only \$102 in quarterly earnings for men and \$141 for women six quarters after random assignment (Bloom et al., 1993); these are roughly equivalent to \$608 in annual earnings for men and \$840 in annual earnings for women in 2010 dollars. Impacts of this magnitude for the current study would be well below the threshold of detection. Indeed, if one were to discount the SEIS results as extraordinarily large relative to historical standards and look instead at how the current MDIs fit into the range between the JTPA and WIA findings—\$600 to \$1,800 for men, \$800 to \$2,600 for women—the current MDIs exceed the upper limit for men and in some sites for women. If the average of midpoints of these ranges, roughly \$1,500, provides a reasonable overall benchmark, an impact of this size for the current programs—whose MDIs

¹⁷ All else equal, and for a given total sample size, a more imbalanced split of the sample between the treatment and the control groups will lead to larger MDIs. The total sample sizes for Grand Rapids and Kern are similar, with Grand Rapids only slightly larger, but Grand Rapids has a more imbalanced distribution between the two research groups. This suggests that the Grand Rapids MDIs will be slightly larger than those for Kern. However, the earnings-based MDI for Grand Rapids is smaller than that for Kern because of the different assumptions for the sites about the percentage of the sample who is male. Kern is expected to have a much higher percentage of males, and the standard deviation of annual earnings of males has been found to be higher than the standard deviation of earnings for females. Therefore, Kern is expected to have a higher standard deviation of annual earnings for the full sample and a larger corresponding MDI for annual earnings.

¹⁸ This study was non-experimental and compared individuals who chose to enroll in WIA training to matched comparison individuals who did not enroll in WIA training. There was evidence that those individuals who chose to participate in WIA training were earning more than their matched comparisons even before entering training. Therefore, these estimates are likely inflated relative to what would be found using an experimental approach.

ranges from \$2,100 to \$3,200—would have far less than an 80 percent chance of detection as statistically significant. While this is not ideal, it represents the limit of the statistical power attainable in a study that seeks to use rigorous random assignment methods to identify one or more ARRA grant-funded training programs proven to increase the long-run earnings of its participants.¹⁹ It also drives home that the evaluation team, and readers of the final study report, should not interpret statistically insignificant effects on earnings as evidence that important gains *did not* occur since even impacts that are fairly large by historical standards—say in the range of \$1,500 to \$2,000 of annual earnings increase—will have less than an 80-percent chance of detection.

6.3 Estimation Methods

The basic impact estimates can be computed using simple subtraction: the difference in the outcomes for treatment group members and control group members is the treatment’s impact. This estimate is unbiased because the individuals who comprise the treatment and control groups were selected at random from a common pool, such that the only systematic difference between the groups likely to show up as statistically significantly different outcomes is that one group had access to the studied training and the other did not. In other words, the simplest test of an intervention impact on some outcome, y , (e.g., earnings) compares the average value of y in the treatment group with the average value of y in the control group. If the difference between these two averages is statistically significant, chance is ruled out as the explanation and we can conclude that the grant programs have an impact. Thus, a random assignment design properly carried out eliminates threats to internal validity due to selection factors, which in non-experimental “comparison group” analyses using naturally occurring program non-participants instead of a randomly assigned control group creates underlying differences between the two groups being compared

While the simple treatment-control differences are unbiased, random differences in the characteristics of treatment and control groups will exist. We will use regression to improve the precision with which we estimate program impacts. The equation that estimates the program’s causal impact, δ , is as follows:

$$y_i = \alpha + \delta T_i + \beta X_i + \varepsilon_i$$

where

y is the outcome of interest (e.g., employment, earnings),

T is treatment status (=1 if randomly assigned to treatment; =0 if control),

δ is the treatment’s impact (the ITT estimator),

α is interpreted as the regression-adjusted outcome for the control group,

X are other control variables measured at baseline,

β reflects the influence of baseline traits, which are not of interest,

ε is the regression residual, and

the subscript i indexes observations.

¹⁹ In particular, in a complete sweep of all grantees participating in the two ARRA grant programs studied, no grantees with larger expected intake flows during the study’s enrollment period were found that were suitable for inclusion in a random assignment study based on the criteria described in Chapter 2.

The primary and confirmatory outcome is earnings, measured as a continuous variable. Exploratory outcomes are both continuous (e.g., weeks worked) and binary (e.g., having a degree or credential). For these we will explore the sensitivity of the results by using a logit model in addition to the standard linear regression model.

For both confirmatory and exploratory analyses, we will conduct two-tailed tests for whether the treatment impact, δ , is statistically significantly different from 0, using a 10 percent standard of evidence but, in the confirmatory analysis, making an adjustment for multiple tests (see discussion in section 6.1 above) of the impact on annual earnings in the four sites. For the exploratory outcomes, we will report 1 percent, 5 percent, and 10 percent thresholds with no adjustments for multiple tests to allow readers—rather than a strict, pre-specified protocol—to decide when the findings are sufficiently suggestive concerning *possible* (not proven) impacts to be relevant to policy. For both confirmatory and exploratory test results, we will report 90-percent confidence intervals to further convey the degree of uncertainty surrounding the findings.

This approach estimates the “intent to treat”, or ITT impacts of the intervention, discussed in section 6.1 above. It shows the average effect of *access* to training, even in instances where the treatment group includes individuals who did not show up for training after being admitted. We are also interested in the average effect of *receiving* training—the effect of “treatment on the treated,” or TOT, as discussed in section 6.1. The conventional approach to estimating the TOT effect is to rescale the ITT estimate—i.e., the overall treatment-control group difference in outcomes for the entire experimental sample—to reflect just those cases that received program services. This methodology assumes that program group members who do not participate have no impact and that control group members who do participate experience the same impact as the equivalent individuals in the treatment group. Computationally, the TOT estimator can be computed as the ITT impact estimate divided by $1 - R_N - R_P$, where R_N is the ARRA-funded training nonparticipation rate in the treatment group (i.e., the no-show rate) and R_P is the ARRA-funded training participation rate in the control group (i.e., the crossover rate).²⁰ Division by this factor, which is always less than 1, increases the estimate of impact (and its standard error); i.e., the effect of participating compared to not participating (TOT) is larger on average than the effect of being granted access to the training through random assignment—in situations where some who are granted access do not participate and/or some of those “denied” access do nonetheless participate. Because the same statistical test for significantly positive (or negative) impact applies to the TOT estimate as to the ITT estimate, no additional adjustment for multiple tests is needed for the annual earnings confirmatory TOT results in the four sites (nor for any of the exploratory TOT findings, of course).

²⁰ See Bloom (2006) for a discussion of this formula (which he characterizes as division by the impact of random assignment on intervention receipt).

Section 7 – Project Activities and Schedule

This section describes the task-by-task activities for carrying out the site selection, design, data collection, analysis, and reporting work for the evaluation. Figure 7-2 shows the timeline for each task. The task schedule has been developed with the assumption that intake will take up to 17 months from its beginning in August 2011. In addition, it assumes that an extension of the evaluation contract will be available to allow for 36-month follow-up surveys to be completed for study participants, including participants who go through random assignment near the end of the intake period, as well as analysis and reporting of the results after the data collection effort is completed. Figure 7-1 lists major deliverables for the evaluation. We will submit all key deliverables electronically to the Department of Labor (DOL) in Microsoft Word, as well as other formats as requested and necessary (such as PowerPoint for presentations).

Figure 7-1. Major Deliverables and Delivery Dates

Major Deliverables (Task Numbers)	Delivery Date
Final Evaluation Design (Task 2)	December 2011
Process Study (Task 3)	In conjunction with the interim and final reports
Monthly Memoranda on the Status of Random Assignment (Task 4)	Ongoing from the start of random assignment (August 2011) to the end of random assignment (December 2012)
Monthly Memoranda on the Survey Sample Size and Follow-up Data Collection (Task 5)	Ongoing from the start of each survey data collection fielding period to the end of each data collection fielding period. For the 18-month follow-up survey, this will be from February 2013 to September 2014. For the 36-month follow-up survey, this will be from August 2014 to March 2016.
Public Use Files and Documentation (Task 5)	September 15, 2016
Interim Report: includes 18-month follow-up and process study (Task 6)	November 30, 2014 (draft) February 28, 2015 (final)
Final Report (Task 6)	June 30, 2016 (draft) September 15, 2016 (final)
Briefings (Task 8)	Up to five briefings, to be scheduled at DOL's request.
Monthly Progress Reports	Ongoing throughout contract, by the tenth of each month
<i>Note:</i> See the text for additional description about the major deliverables shown in this table, as well as other reports and information that we will provide to DOL throughout the evaluation.	

7.1 Task 1: Selection of Sites

The first task, selection of sites for inclusion in the study, is complete. We examined the universe of grantees through a systematic review of their applications and, in conjunction with DOL, decided to focus the study on grantees that were both implementing a service model that targets career pathway programs and offering comprehensive supportive services. We conducted phone interviews with 30 grantees, and ten grantees were then selected for in-person visits, conducted in March 2011. Based on the information from these visits and DOL's knowledge of the grantees, the four grantees that are included in the study were notified of their selection for the study in April 2011.

7.2 Task 2: Evaluation Design

Task 2 consists of the development of the study design and the submission of the OMB package to obtain clearance for the data collection forms to be used as part of the intake of participants into the study. The design task has included the creation of a conceptual framework for the study, a determination of the numbers of treatment and control group members necessary at each study site to answer the research questions, development of random assignment procedures and implementation plans that were tailored to each site, and the design of baseline data collection instruments and procedures. It culminates in the current document, the study's *Evaluation Design Report*.

In addition, the design task includes submission of an OMB clearance package that covers the consent form and the baseline data collection form as well as related work to obtain the clearance. Emergency OMB clearance was requested in April 2011, so that as many grant participants as possible could go through the random assignment process before the grants expire. Emergency clearance was received in July 2011 and is being followed by a request for regular OMB clearance of the baseline forms as well as the process study site visit data collection forms. (OMB clearance for the follow-up surveys is described under Task 5 below.)

7.3 Task 3: Process Study

Task 3, the process study, includes two rounds of data collection, analysis, and reporting. Both rounds of data collection and analysis will cover the implementation experiences of all four sites included in the evaluation. Site visitors will use prepared discussion guides to conduct the semi-structured interviews. They will be guided by a set of protocols that cover the types of information required to advance our understanding of the training programs. We will also hold a site visitor training for all staff involved in conducting the visits.

The visits will consist of interviews with program staff, partners and employers; focus groups; and observations of grantees' activities. The first round of visits will be conducted approximately six months after the start of random assignment (in February 2012, pending receipt of OMB clearance for the site visit data collection effort) and the second will occur 14 to 18 months after the start of random assignment (in October 2012 through February 2013²¹). Detailed findings from the process study will be presented in the interim report (to be provided in February 2015). The final report, to be provided in September 2016,

²¹ Because the Pathways out of Poverty grants were established for a shorter time frame than was the case for the Health Care grants, the Grand Rapids second-round site visit will likely be scheduled earlier, to allow us to meet with staff prior to the end of the grant.

will include a summary of these results, with a focus on the process study results that inform the final impact analysis and overall findings from the evaluation.

7.4 Task 4: Implementation and Monitoring of Random Assignment

This task includes the preparation of site-specific materials for implementation of the study at each site, the training of site staff, and collaboration with and monitoring of sites during the random assignment period to ensure that study implementation is successful. Planning for random assignment at each site occurred in May through August 2011 and involved an assessment of the projected number and flow of study participants during the random assignment period, identification of strategies to ensure that an adequate number of study participants can be recruited for the study, and the tailoring of generic study procedures to each site's unique grant features. Furthermore, we worked with site staff, as needed, to ensure that barriers to participation were addressed.

Random assignment at each study site began in August 2011, immediately after the training of site staff on study procedures. Evaluation staff was present in person at each site at the start of random assignment, and we will have frequent contact with site staff throughout the random assignment period, which is expected to last from August 2011 to December 2012. Because we expect that staff will have more questions about study procedures during the first few months of the random assignment period, we will contact site staff weekly during this period. However, at later points, we will still contact sites at least on a monthly basis to ensure both that there is adherence to study procedures and that challenges encountered by site staff are addressed quickly. As part of our monitoring of sites, we will conduct reviews of (1) data in the Participant Tracking System, which contains baseline data about study participants; and (2) documentation of calls between sites and their liaisons to the research team.

Throughout the random assignment period, we will provide monthly reports to DOL about the build-up of the number of study participants, as well as any problems that the sites have encountered that might influence their ability to achieve the target sample sizes. Based on our assessment of each site's situation, we will work with sites (and DOL, as needed) to develop and implement strategies to mitigate or avoid the problems.

7.5 Task 5: Follow-Up Surveys

This task consists of several activities related to the 18- and 36-month follow-up surveys: obtaining OMB clearance for the surveys; preparing for and fielding the surveys; keeping DOL informed about the progress of the surveys; and cleaning of the data after the data collection efforts are complete. The task also includes the preparation and submission of restricted and public use data files to DOL.

As with the intake forms, OMB must approve the survey data collections activities before those activities can begin. DOL must submit a notice for publication in the *Federal Register* describing the proposed data collection, with instructions on how the public can comment on the proposed data collection. The *Federal Register* notice provides a brief description of the study goals, the types of data to be collected, the respondents, and the total burden of the data collection on respondents. We will prepare a draft of this notice and submit it to DOL. We will revise the notice, as needed, in response to DOL comments and resubmit it. At least 60 days must be allowed for public comments, which have to be addressed in the OMB package, if any are received.

The OMB package will include drafts of the 18- and 36-month instruments developed for the project and a supporting statement that includes a careful description of the study goals, sample design, burden on

respondents, and analysis plans. The package will describe our strategies to achieve high response rates and our justification for the \$25 incentive payments per survey that we plan to provide to respondents. We will submit the draft OMB package to DOL in January 2012. We will then revise the package in response to comments received from DOL, or the public in response to the published notice, June 2012. We anticipate that DOL will submit the package for OMB clearance in October 2012, and we will receive OMB clearance in January 2013, prior to the start of 18-month follow-up interviewing in February 2013.

The data collection work for the 18- and 36-month surveys includes the fielding of the surveys, but also activities to be conducted prior to and after the fielding period. For each survey, we will program the survey instrument so that it can be administered using computer-assisted telephone interviewing (CATI) technology. The survey-related effort also includes interviewer training; attempts to locate sample members and to address the concerns of sample members who have initially declined to participate; and maintenance of a centralized data management system to track progress, monitor quality, and link data files. The fielding of the 18-month follow-up survey is expected to occur from February 2013 to September 2014, while the comparable time frame for the 36-month survey is from August 2014 to March 2016. We expect to attempt to locate and administer each survey to all 4,000-plus study participants in each wave, achieving an 80 percent response rate and around 3,200 completed interviews at the two time points. After each fielding period is completed, we will conduct data cleaning activities to ensure that the data files are ready for analysis.

For the fielding of both the 18- and 36-month surveys, we will provide monthly reports to DOL about the number of completed surveys and the response rate. The reports will also describe any problems that we have encountered during the data collection effort, as well as our proposed solutions.

Near the end of the project, we will prepare both restricted and public use data files and documentation. Both will include participant-level baseline and follow-up data, data dictionaries, and a user's guide. To ensure that individual study participants cannot be identified, we will remove, mask, or encrypt identifying information. For each variable, the data dictionary will include its name, a description, its source, a frequency distribution, and other information that could be pertinent to users of the data file. Both raw and key analysis variables will be included. Restricted use files, which may contain some identifiable information, will only be available to individuals after they have submitted an application and received approval from DOL to use the data. As part of this process, we understand that DOL will require investigators to sign a confidentiality agreement.

The data files will be provided in ASCII, with a machine-readable file layout showing the position of each data element. We will also provide instruction programs that will allow for quick conversion of the data to SAS, STATA and SPSS formats and for use with other applications like Excel. We will submit the public use data files to DOL in September 2016.

7.6 Task 6: Preparation of Reports

Task 6 consists of analysis and reporting for both the interim and final reports. The interim analysis and reporting activity will provide to DOL information about the early experiences and outcomes of study participants, while the final analysis and reporting will provide a longer-term perspective on the impacts of study participants, including the impact on our primary confirmatory outcome measure of earnings.

Our schedule will ensure that we can provide findings for the interim and final analyses in a timely manner. Since the interim report will include analysis of the baseline data, as well as findings from the process study, we will be able to develop chapters of the report while we await the completion of the 18-

month follow-up survey. In addition, we will begin the analysis of survey data using a preliminary extract of the data. For the interim report, we will begin the analysis in June 2014, prior to the completion of the 18-month survey fielding period. Although the preliminary survey file will need updating after all survey data become available, we expect to have at least three-quarters, and possibly more, of the completed surveys in the preliminary file. This strategy will allow us to develop and revise our computer programs to estimate impacts, and to identify and follow up on additional lines of inquiry, while we await the end of the survey fielding period. Furthermore, we will begin working with the administrative data at the end of June 2014.²² Because administrative data contain few data items and are generally fairly clean files, we expect that the processing of these data will be straightforward and that we will be able to quickly adapt the survey-focused computer programs that estimate program impacts.

An outline of the interim report will be submitted in August 2014, 90 days before the draft report is due in November 2014. We will submit the final version of the report in February 2015, no later than two months after we expect to receive comments from DOL. In addition to the electronic version of the report, we will provide 10 paper copies to DOL. Although the outline of the report will be decided in collaboration with ETA, we expect at this point that the report will contain two sections: a concise main report (about 30 to 50 pages in length) and a set of technical appendices, which will provide details about the study's random assignment procedures and analysis methodology. We also will provide DOL with a half-page summary or "research brief" that highlights the key findings of the study, as well as a stand-alone executive summary of the report (approximately 5 pages).

The activities and work flow for the final report will be very similar to those for the interim report, except the focus will be on the data collected at the 36-month follow-up period. As with the interim analysis, we also will use a preliminary extract of the survey data, made available before the completion of the survey fielding period. We will begin this analysis in December 2015, and we will be able to work with the administrative data after we receive the data extracts from states at the end of the same month.²³ We expect to be able to adapt many of the computer programs developed during the interim report analysis for use with the final report analysis. A final survey data file will be available at the end of March 2016, allowing for three months to update estimated results and draft text before the draft of the final report will be provided to DOL in June 2016. An outline of the report will be submitted in March 2016, 90 days before the draft report is due. As with the interim report, the final version of the report will be submitted no later than two months after we receive comments from DOL, and we will provide 10 paper copies to DOL in addition to the electronic version of the report.

We envision a final report with a structure similar to that of the interim report. As for the interim report, we expect to provide a main section of the report, technical appendices, a half-page summary of the study, and a stand-alone executive summary.

²² As discussed in Section 6, the schedule for the evaluation is unlikely to allow us to collect a full 18-month follow-up period for sample members who go through random assignment near the end of the sample member intake period. As we negotiate with the states that will provide us with administrative data extracts, we will be clear when we need the data so that we will be able to adhere to the evaluation's timeline for providing the interim report to DOL.

²³ As with the 18-month administrative data collection effort, the schedule for the evaluation does not allow for us to collect full 36-month follow-up data for sample members who enroll in the study near the end of the intake period. We will work with states to ensure that the timeline for their provision of the data is feasible for them and will allow us to provide the final report to DOL on schedule.

7.7 Task 7: Peer Review Panel

Over the course of the project, the PRP will provide feedback on key project documents. The PRP consists of five research and program experts, who will be convened three times for in-person meetings. Each in-person meeting will last a day and will be held in or near Washington, D.C. The first meeting, held in July 2011, was to solicit input on the planned evaluation design, methods, data analysis, and intake data collection forms presented in this design report. The second and third meetings will provide an opportunity for us to receive the PRP's input on the analysis and draft deliverables arising from the 18- and 36-month follow-up data collection. At least three weeks prior to each meeting, we will send a proposed agenda and written background materials to DOL for approval; after receipt of approval we will send this information to the PRP members, which they can review in advance to facilitate a smooth and efficient meeting. Within two weeks after each meeting, we will submit to DOL a memo summarizing the meeting.

The PRP members were selected to represent a range of expertise in areas of importance to the study. Collectively, they have expertise in the following areas: (1) training of workers with low skill and/or low income levels; (2) the health care and green jobs labor markets; (3) random assignment evaluation design and analysis; and (4) survey methods.

7.8 Task 8: Oral Briefings

At DOL's request, we will present up to five oral briefings on the evaluation. Topics could cover the evaluation design, the interim data analysis and report, and the final data analysis and report. In Figure 7-1, we have shown that these briefings will be spread throughout the evaluation period.

7.9 Task 9: Administrative Data Collection and Protection of Personally Identifiable Information

The study design calls for two rounds of administrative wage records data, which will be collected from either the four states that contain the study sites or via the NDNH. To maximize the likelihood that states will provide the data (should that avenue be pursued), we will work closely with DOL to develop and finalize advance information about our request to the states. We then will send the advance information to the states, contact each state by telephone to confirm its participation, and (as needed) negotiate memoranda of understanding. The memoranda of understanding will include the following parties: (1) a state, (2) Abt Associates, (3) Mathematica, and (4) the DOL, as needed.

We will request that states provide the wage records data in two extracts; doing so is intended to serve two purposes. First, it will minimize burden on the states, given that some states archive their wage data records after about five years has passed since the quarter to which the data pertain. Second, the availability of two data extracts will allow us to examine administrative measures of employment and earnings as part of both the interim and final analyses for the study. (Because of a lag in the reporting of earnings information by employers to state agencies, and the time that it will take for states to process the data and provide us with data extracts, the first extract may cover about 12 months of the follow-up period, rather than 18 months, for the study participants who went through random assignment near the end of the follow-up period. In a similar way, the second extract may cover about 30 of the 36 months for those study participants.) We expect to receive the first set of data extracts by the end of June 2014 and the second set by the end of December 2015.

Administrative data provided by UI agencies will be included in the public and restricted use files and supporting documentation that will be provided to DOL.²⁴ The administrative data will undergo the same procedures to ensure that personally identifiable information will be masked or removed from the files. These public use files will follow the current OMB checklist on confidentiality to ensure that the file and documentation can be distributed to the general public for analysis. Steps will be taken to ensure that sample members cannot be identified in indirect ways. For example, categories of a variable will be combined to remove the possibility of identification due to a respondent being one of a small group of people with a specific attribute. Variables will also be combined in order to provide summary measures to mask what otherwise would be identifiable information, and some continuous variables (such as earnings) may be converted to discrete variables. These strategies might be especially likely to be used for administrative data, to prevent the risk of identification of sample members. Although it cannot be predicted which variables will have too few respondents in a category, we plan not to report categories or responses that are based on cell sizes of less than five. If necessary, statistical methods will be used to add random variation within variables that would be otherwise impossible to mask.

7.10 Monthly Progress Reports and Management

As part of our ongoing monitoring of the project and communication with DOL staff, we will conduct management activities on a monthly basis, and more frequently as needed, to examine the status of the project's budget, schedule, and technical situation. These activities include submission of progress reports to DOL by the tenth day of each month. Each report will contain a summary of work completed during the previous month, work planned for the upcoming month, problems encountered and proposed solutions, and outstanding needs from DOL.

²⁴ If quarterly earnings data from the NDNH are used in lieu of state-supplied wage records data, it will not be possible to include those data in the public and restricted use files because of restrictions on data confidentiality.

Works Cited

- Benjamini, Yoav and Yosef Hochberg (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society, Series B (Methodological)*, 57(1): 289-300.
- Bloom, Howard S. (1984). "Accounting for No-shows in Experimental Evaluation Designs." *Evaluation Review*, 8(2): 225-246.
- Bloom, Howard et al. (1993). "The National JTPA Study: Title II-A Impacts on Earnings and Employment at 18 Months." Bethesda, MD: Abt Associates, Inc.
- Bloom, Howard S. (2005). "Randomizing Groups to Evaluate Placed Based Programs" in Learning More from Social Experiments: Evolving Analytic Approaches, Howard S. Bloom (ed)., Ch.4, pp.115-172.
- Bloom, Howard (2006). "The Core Analytics of Randomized Experiments for Social Research." MDRC Working Papers on Research Methodology. New York, NY: MDRC. Available at <http://www.mdrc.org/publications/437/full.pdf>
- Dohm, Arlene and Lynn Shniper. (2007). "Occupational Employment Projections to 2016." *Monthly Labor Review*, 130(11): 86-125.
- Greenberg, David H., Charles Michalopoulos, and Philip K. Robins. (2006). "A meta-analysis of government-sponsored training programs." *Industrial and Labor Relations Review*, 57(1): 31-53.
- Greenberg, David H., Charles Michalopoulos, and Philip K. Robins. (2004). "What Happens to the Effects of Government-Funded Training Programs Over Time?" *Journal of Human Resources*, 39(1): 277-293.
- Heinrich, Carolyn, Peter R. Mueser, and Kenneth R. Troske. (2009). "Workforce Investment Act Non-Experimental Net Impact Evaluation: Final Report." Washington, D.C.: U.S. Department of Labor, Employment and Training Administration Occasional Paper 2009-10.
- Hotz, V.J., Imbens, G. W. & Klerman, J.A. (2006). "Evaluating the differential effects of alternative welfare-to-work training components: A re-analysis of the California GAIN Program." *Journal of Labor Economics*, 24 (2), 521-566.
- Hotz, V. Joseph and John Karl Scholz. (2009). "Measuring Employment Income for Low-Income Populations with Administrative and Survey Data." Washington, DC: U.S. Department of Health and Human Services, Assistant Secretary for Planning and Evaluation.
- Kluge, Jochen. (2010). "The effectiveness of European active labor market programs." *Labour Economics*, 17(6): 904-918.
- Kornfeld, Robert, and H.S. Bloom. (1999). "Measuring Program Impacts on Earnings and Employment: Do Unemployment Insurance Wage Records from Employers Agree with Surveys of Individuals?" *Journal of Labor Economics*, 17(1): 168-197.

- Lechner, Michael, and Conny Wunsch. (2006). "Are Training Programs More Effective When Unemployment is High?" Swiss Institute International Economics and Applied Economic Research for Working Paper.
- Maguire, Sheila, Joshua Freely, Carol Clymer, Marueen Conway, and Deena Schwartz. (2010). "Tuning Into Local Labor Markets: Findings from the Sectoral Employment Impact Study." Philadelphia, PA: Public/Private Ventures.
- McConnell, Sheena. (2006). "Managing Customers' Training Choices: Findings from the Individual Training Account Experiment." Washington, DC: Mathematica Policy Research.
- McConnell, Sheena M., Elizabeth A. Stuart, Kenneth N. Fortson, Paul T. Decker, Irma L. Perez-Johnson, Barbara D. Harris, and Jeffrey Salzman. "Managing Customers' Training Choices: Findings from the Individual Training Account Experiment." (2006). Final report submitted to the U.S. Department of Labor, Employment and Training Administration. Washington, DC: Mathematica Policy Research.
- Mills, Gregory, Daniel Gubits, Larry Orr, David Long, Judith Feins, Bulbul Kaul, Michelle Wood, Amy Jones and Associates, Cloudburst Consulting, and the QED Group. (2006). Effects of Housing Vouchers on Welfare Families: Final Report. Prepared for the U.S. Department of Housing and Urban Development, Office of Policy Development and Research. Cambridge, MA: Abt Associates Inc.
- Schochet, Peter Z. (2008). Technical Methods Report: Guidelines for Multiple Testing in Experimental Evaluations of Educational Interventions. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Available at http://ies.ed.gov/ncee/tech_methods/
- Schochet, Peter Z. (2009). "An Approach for Addressing the Multiple Testing Problem in Social Policy Impact Evaluations." *Education Review* 33(6): 9-149. Available at <http://erx.sagepub.com/content/33/6/539.full.pdf+html>
- Schochet, Peter Z., Sheena M. McConnell, and John A. Burghardt. (2003). "National Job Corps Study: Findings Using Administrative Earnings Records Data." Report submitted to the U.S. Department of Labor, Employment and Training Administration, Office of Policy and Research. Princeton, NJ: Mathematica Policy Research.
- U.S. Department of Labor (2004). "Comparison of State Unemployment Insurance Laws." U.S. Department of Labor: Employment and Training Administration Office of Workforce Security. Washington, D.C. Available at <http://workforcesecurity.doleta.gov/unemploy/comparison.asp>.