Green Jobs and
Health Care
Impact Evaluation

OMB Clearance Package for Follow-Up Data Collection

Part B

January 2012

CONTENTS

TABLES

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

The Employment and Training Administration (ETA) in the U.S. Department of Labor (DOL) is undertaking the Green Jobs and Health Care (GJ-HC) Impact Evaluation of the Pathways Out of Poverty and Health Care and Other High Growth and Emerging Industries Training grant initiatives. The goal of this evaluation is to determine the extent to which enrollees achieve increases in employment, earnings, and career advancement as a result of their participation in the training provided by Pathways and Health Care grantees and to identify promising practices and strategies for replication. ETA has contracted with Abt Associates and its subcontractor, Mathematica Policy Research, to conduct this evaluation.

On July 18, 2011, ETA received Office of Management and Budget (OMB) approval[1] for the baseline data collection effort. A separate submission for the process study data collection, which includes site visits and focus group administration, has been submitted and is under review. The request for clearance included in this package is limited to the follow-up interviews to be attempted with all study participants 18 months and 36 months after baseline collection (questionnaire presented in Attachment 1).

The full package for this evaluation needed to be submitted in three parts for several reasons. The main reason is that it was necessary to (1) conduct random assignment and collect baseline data early in the study period in order to obtain a sample size needed for the estimation of program impacts and (2) conduct two rounds of process study visits, including one when the early sample was participating in the training program. In addition, the study structure required that the baseline data inform the development of the follow-up data collection effort. As a result, it was necessary to first obtain clearance for the baseline data collection and gain experience in its implementation before the follow-up instruments could be developed and then submitted for clearance.

A. Respondent Universe and Sampling

The evaluation will measure the effectiveness of the training strategies adopted by four grantees selected from among the 93 grantees funded under the Pathways Out of Poverty and the Health Care and Other High Growth and Emerging Industries programs. The evaluation team based selection of these grantees primarily on the strength and scale of the grantees' intervention and their ability to support the requirements of this type of evaluation. The study will not indicate whether the two grant funding vehicles as a whole produce beneficial effects, but rather it will tell ETA

---

[1] OMB assigned 1205-0481 as the OMB Control Number for the baseline data collection effort.

whether any of the specific training approaches used in the four study sites are worth emulating as successful models for improving workforce outcomes in the green jobs and/or health care sectors.

At the four sites that are included in the study there is no sampling to identify study participants; instead, this study is a census of all eligible applicants to the grant-funded program at each study site (who also consent to participate in the study).   Using a census is necessary because the population for each site is not large; a site-specific sample size that is much smaller than the population (such as one generated through random sampling of study participants) is less likely than the full site population to generate a statistically significant impact estimate of a magnitude relevant to policymakers.   In addition, given the differences in the treatment interventions across sites, it is not appropriate to pool the participants in the four sites to conduct the impact analysis.. Site staff use their existing eligibility criteria to identify people who are qualified to receive program services.   All of those individuals are asked to complete an Informed Consent form and a Baseline Information Form before they can be considered for grant-funded services and be randomly assigned into either a treatment group that has access to those services or a statistically equivalent control group that does not. A total of 4,024 sample members (of which 2,214 will access grant-funded services) are expected across the four sites, with target sample size totals varying by site as shown in Table B.1.

As with the baseline data collection effort, all the individuals randomly assigned to either the treatment or control group will be contacted for inclusion in each of the 18-month and 36-month follow-up telephone surveys; statistical methods will not be used to select a subsample. The treatment: control ratios were determined as part of negotiations within each of the four study sites. The main considerations were the number of potential people that each site expected to be able to serve after random assignment, along with the site's service number targets as agreed upon with DOL.

Table B.1. Sample Size Requirements

Number of Participating Sites		4			
	Site	Treatment	Control	Approximate Treatment-Control Ratio	Total
	AIOIC (MN)	600	600	1:1	1,200
Projected number of treatment and control group members/ ratio of treatment to control group members	Grand Rapids (MI)	600	300	2:1	900
	North Central Texas	589	485	12:10	1,074
	Kern (CA)	425	425	1:1	850
	All sites	2,214	1,810	---	4,024
Anticipated number of respondents to 18 month	AIOIC (MN)	480	480	1:1	960
	Grand Rapids (MI)	480	240	2:1	720

	Site	Treatment	Control	Approximate Treatment-Control Ratio	Total
Number of Participating Sites	4				
	North Central Texas	471	388	12:10	859
	Kern (CA)	340	340	1:1	680
	All sites	1,771	1,448	---	3,219

*Assuming an 80 percent response rate at each point in time.

B. Analysis Methods and Degree of Accuracy

1. Statistical Methodology for Stratification and Sample Selection

None of the data collection activities for the GJ-HC Impact Evaluation requires a statistical methodology for stratification or sample selection given that the evaluation is a census of participants within the four study sites and that no sampling of the study sites was conducted.

2. Estimation Procedure

As described in Supporting Statement Part A, Section A.2, the primary objective of the evaluation is to estimate program impacts—that is, observed outcomes for the treatment group relative to what those outcomes would have been in the absence of the program—in each of the four programs studied. Specifically, the study will identify the extent to which grant-funded training and services in a given site improve participant employment, earnings, and career advancement. The study will also attempt to identify promising training practices and potential strategies for replication of successful interventions. This section lays out the research questions for the evaluation, describes our analytic approach to estimate program impacts using the 18- and 36-month survey data that are included in this request for clearance, 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 (that is, the minimum detectable impacts).

As noted in Part A, the two follow-up surveys submitted with this package look at outcomes for the treatment and control group members. The evaluation will address the following research questions:

- What is the impact of the selected grantee programs on the receipt of education and training services by treatment group members, in terms of both the number who receive these services and the total hours of training received?

- What is the impact of the programs on the completion of training and educational programs and on the receipt of certificates and credentials from these programs?

- What is the impact of the programs on employment levels and earnings? To what extent do the programs result in earnings progression?

- To what extent do the programs result in any employment (regardless of sector)? To what extent do the programs result in employment in the specified sector in which the training was focused?

- What features of the programs seem to be associated with positive impacts, particularly in terms of target group, curricula and course design, and additional supports?

- What are the lessons for future programs and practices?

The basic impact estimates can be computed using simple subtraction: the difference in average outcomes between the treatment group members and control group members in that site is an unbiased measure of the impact of having access to the intervention. Each of the resulting four estimates for each outcome is unbiased because the individuals who comprise the treatment and control groups in the site were selected at random from a common pool and hence are statistically equivalent on all factors at baseline, in expectation. As a result, any statistically significant differences in outcomes between the groups can be attributed to the effects of the intervention. In other words, the test of an intervention impact on some outcome, *y,* (for example, earnings) in a site 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 significantly different from zero, chance is ruled out as the explanation and we can conclude that the grantee's program has an impact on the measured outcome. Thus, random assignment properly carried out eliminates threats to internal validity due to selection into the treatment group and other factors.[2] This is different from a non-experimental comparison group analysis—that is, using naturally occurring program nonparticipants instead of a randomly assigned control group—in which underlying differences between the two groups being compared remain even with statistical adjustment, leading to potential problems with internal validity of the estimates.

While the simple treatment-control differences are unbiased, random differences in the characteristics of treatment and control groups will exist, increasing the variance of impact estimates. Regression analysis that controls for variations in measured background characteristics between individuals will be used to improve the statistical precision of the impact estimates. The specific implementation approach for these controls will vary with the outcome. A standard linear regression will be used for continuous outcomes like earnings. A logistic model will be used for binary outcomes, such as having a degree or credential.

---

[2] Impact results for a particular site will not be generalized beyond that site.

We will construct weights to be used in the analysis. Because neither the sites nor the study participants are selected with probability methods, we will not construct a sampling weight to account for the probability of selection, nor do we plan to generalize our findings beyond each site and the set of study-eligible individuals who agreed to participate in the study. We will, however, construct randomization weights that account for each participant's probability of selection into the treatment or control group. Each randomization group is essentially a random sample of all participants in each site.

We will then adjust these randomization weights to account for nonresponse within site and randomization group in an effort to minimize the risk of nonresponse bias. Within each site and group we will conduct a nonresponse analysis using baseline and demographic information collected on all study participants in order to determine which characteristics are both associated with a propensity to respond and correlated with the key outcomes being measured. We will then run response propensity models using these predictive variables and use the resulting propensity score to form weighting cells within site and group.

3. Degree of Accuracy Needed for the Purpose Described in the Justification

As described in Section B.1, an anticipated 4,024 study participants will result in approximately 3,219 completed 18-month telephone surveys assuming an 80 percent completion rate. Eighteen months later, the 36-month telephone interviewers will attempt to contact and interview the entire group of 4,024 participants. Again, assuming an 80 percent completion rate, approximately 3,219 sample members will complete the 36-month interview. Because all impact analyses will be conducted at the site level, given the existence of important differences between the interventions at the four sites, the site-specific sample sizes are the sizes that are most important. The site-specific sample sizes will provide adequate precision, expressed in terms of minimum detectable impacts (MDI), to support the evaluation.

MDIs are the smallest true impacts that the study has at least an 80-percent probability of detecting as statistically significant; for a given level of power, the greater the sample size, the smaller the MDI that can be detected. It is important to calculate MDIs before beginning an evaluation to ensure that the study will be able to detect impacts of magnitudes that are relevant to policymakers. In this section, MDIs for three representative outcomes of interest are presented, given the projections of likely sample sizes at each grantee.

MDIs are a function of several factors, among them 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, and not the sum across grantees, so we do not need to account for within-site clustering effects on the variance. We do need to account for the effect of nonresponse-adjusted weights on the variance of estimates. Using specialized procedures designed for analyzing survey data (within SAS or Stata, or using SUDAAN, and utilizing the Taylor Series Linearization method), the calculation of the variance for each estimate will appropriately account for design effect due to unequal weighting. Because we do not expect large variation in response rates within site and randomization group, this design effect should be minimal. The study does not seek to determine whether the two grant funding vehicles as a whole produce beneficial effects for society. Rather, it is motivated by a desire to discover whether any of the specific training approaches used in the four study sites are worth emulating as successful models for improving workforce outcomes in the green jobs and/or health care sectors. Hence, all analyses will be conducted separately by site as four independent tests of specific training interventions.

A litmus test for the effectiveness of each intervention will be its impact on participant earnings in months 25 to 36 of the follow-up period. The evaluation will focus on a single outcome for this litmus test (what is referred to in the literature as a confirmatory outcome), rather than a broader set of outcomes, to reduce the extent of a problem that arises when more than one statistical test is conducted. The problem is that, even when 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 conducted. Because the evaluation will conduct four tests on the confirmatory outcome—one for each of the four study sites—a multiple comparisons adjustment will be made to standard

hypothesis test procedures.[3] The MDI calculations take this adjustment into account.

In light of this decision, an overall Type 1 error rate of 10 percent (alpha = .10) has been adopted rather than a 5 percent rate (alpha = .05). This is because a five percent error rate—when combined with the multiple comparisons adjustment to take into account the impact estimation for four sites—would tilt testing of the impact estimates too much in the direction of avoiding false positive results (Type 1 error) while unduly expanding the risk of false negatives (that is, a finding of no significant impact in any site when an effect of important magnitude has in fact occurred—Type 2 error).

The first two columns of Table B.2 present the projected treatment and control group sample sizes for each of the four study sites. These projections assume a 12- to 18-month random assignment sample intake period in each site as currently projected. The third column with numbers presents the MDIs for the primary confirmatory outcome in each site, total earnings in months 25 to 36 after random assignment. The final two columns of the figure present MDIs for two illustrative exploratory outcomes of interest— analyses that seek to explore suggestive evidence of possible impacts rather than conclusive evidence of effects through confirmatory analysis such as will be conducted for earnings. The particular illustrative exploratory outcomes considered are employment in month 36 after random assignment and possession of a degree or credential in that month. We do not plan to apply a multiple comparisons adjustment in the exploratory analyses given the more tentative nature of the conclusions that will be drawn from them.

---

[3] A goal of the evaluation is to have evidence that conclusively *confirms* that earnings impacts have occurred, rather than that—as with the other impact measures to be examined —simply *explore* whether effects may have taken place. (See Schochet 2009 for a discussion of the difference between confirmatory and exploratory tests of intervention effects, and Abt Associates 2011 for application of these concepts to the GJ-HC Impact Evaluation.)   This requires rigorous control over the Type 1 error probability when conducting multiple confirmatory tests. Suppose, for example, that each of four tests is set up to have a 10 percent chance of rejecting the null hypothesis that earnings in a particular site were not increased by the intervention (that is, there is an alpha of .10). Then, even if none of the interventions increased earnings, the chance of one or more significant findings from the four tests rises to 34 percent (= 1 minus [.9 to the fourth power], or .34). A large literature exists regarding the best way to protect against this inflated risk of a false positive conclusion when doing multiple tests of potential impacts. Calculations of the MDIs for this submission are based on a conservative assumption that the Bonferroni adjustment will be applied to the four earnings impact tests since any procedure that might eventually be used to test the confirmatory impact hypotheses will achieve at least that small of MDIs for the sites. The Bonferroni method reduces the cutoff for a significant effect in each individual test from a p-value below .10 to a p-value below .025, so that the combined probability of Type 1 error across four tests, bounded above by .025 x 4, cannot exceed .10. This adjustment procedure is conservative and has been improved upon in the literature. The adjustment procedure to be applied during the impact analysis will depend on developments in the literature as statisticians continue to find ways to increase statistical power (that is, reduce the attainable MDIs) when testing multiple hypotheses.

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 varies across sites as shown in Table B.1 above) is maintained as constant throughout the sample intake period in any given site.

- The follow-up surveys used to measure outcomes for the impact analysis will achieve an 80 percent response rate,[4] resulting in the target sample sizes shown in Table B.1.

- Two-tailed statistical tests are conducted.

- 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.)[5] 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 percent for AIOIC, 40 percent for Grand Rapids, 20 percent for North Central Texas, and 95 percent for Kern.

- The share of the control group employed in the 36th month after random assignment is 65 percent.[6]

- The share of the control group with an educational degree or training credential in the 36th month after random assignment is 30 percent.[7]

- The inclusion of baseline characteristics of sample members as covariates in the impact regressions will account for approximately 20 percent of the total variation in individual outcomes.[8]

---

[4] Although administrative data with 100 percent coverage for the annual earnings and employment rate outcomes could be used for some of the impact analysis, the conservative approach presented here assumes 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.

[5] 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).

[6] This employment rate is based on the results of the National JTPA Study.

[7] 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 percent 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.)

[8] 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

For the confirmatory outcome, annual earnings in months 25-36, a Bonferroni 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—that is, statistically most powerful—methodology available in the literature when the first impact analysis is conducted; therefore, the true MDIs for annual earnings are likely to be smaller than those shown here.

The MDI formula used for the calculations is as follows:

$$MDI = factor * \sigma \sqrt{\frac{(1-R^2)*1}{nP(1-P)}}$$

As explained above, all MDI calculations assume two-tailed tests and 80 percent power. A 10 percent significance level is used when adjusting for multiple tests of impacts on earnings in the four sites and a .05 significance level used individually in each site when testing impacts on employment rate and degree or credential.. σ is the standard deviation of the outcome, R^2 is assumed to be 0.20, n is the number of survey respondents and P and 1-P are the proportions of respondents allocated to the treatment and control groups, respectively.[9] .   The MDI calculations for annual earnings are adjusted for multiple testing using the Bonferroni approach, whereas the MDI calculations for other outcomes are not. Therefore, the MDI translation factor for the impact estimate standard error equals 3.09 for earnings and 2.80 for the other outcomes.

The standard deviation for the earnings outcome is assumed to vary across sites because (1) the standard deviation of earnings for men is different from that for women and (2) the expected proportions of sample members who are male and female differ across sites.[10] The earnings standard deviation is calculated using the following equations:

---

low-income adults, reported explanatory power of 14 to 19 percent, while the National Job Corps Study achieved 20 percent on this measure.

[9] The formula above is given in Bloom (2006) ["The Core Analytics of Randomized Experiments for Social Research".](#)

[10] A common standard deviation is assumed in all sites for employment and educational attainment.

$$\sigma = \sqrt{\left(\frac{n_m + n_f}{n_m + n_f - 1}\right) * \left(\left(percent_m * term_m\right) + \left(percent_f * term_f\right)\right)}$$

and

$$term_m = \left(\left(\frac{n_m - 1}{n_m}\right) * \sigma_m^2\right) + \left(earnings_m - earnings_{full}\right)^2$$

.

A comparable equation for $term_f$ was used for females. In the equations, n_m and n_f are the numbers of sample members who are male and female, respectively; $percent_m$ and $percent_f$ are the proportion of sample members who are male and female, respectively (such as 0.2 and 0.8 for AIOIC). σ_m^2 is the variance of the earnings of men, $earnings_m$ is the average earnings of men, and $earnings_{full}$ is the average earnings for the full sample (of men and women).

Table B.2. MDIs for Confirmatory and Exploratory Outcomes of Interest

	Treatment Survey Responses	Control Survey Responses	MDI: Annual Earnings, from Months 25 through 36 of Followup	MDI: Employment Rate	MDI: Degree or Credential
AIOIC (MN)	480	480	$2,188	7.7%	7.4%
Grand Rapids (MI)	480	240	$2,920	9.4%	9.1%
North Central Texas	471	388	$2,323	8.2%	7.9%
Kern (CA)	340	340	$3,351	9.2%	8.8%

Note: The site-specific sample sizes are based on an assumed 80% response rate to the 36-month follow-up survey. These sample sizes are used to calculate each of the site-specific MDIs. The MDIs for annual earnings take into account use of a Bonferroni multiple comparison adjustment, given that annual earnings is the primary confirmatory outcome. The calculations for the MDIs for the employment rate and the possession of a degree or credential do not include a multiple comparison adjustment, given that these outcomes are considered to be exploratory. See the text for additional discussion of the assumptions underlying these MDI calculations.

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 under-skilled, unemployed, and low-income adults, found impacts on earnings in the second year of followup of $3,777 for men and $4,555 for women (Maguire et al. 2010). Another study, the Workforce Investment Act Non-Experimental Net Impact Evaluation (Heinrich et al. 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.[11] Although neither of these studies is 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.[12]

It is important to note that one very prominent—but now very dated—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.

C. Methods to Maximize Response Rates and Data Reliability

All study participants, including both treatment and control group members, will be contacted to complete the first telephone survey 18 months after their random assignment date and the second survey 36 months after random assignment. The expected response rates are 80 percent for 18-month telephone interview and 80 percent for the 36-month

---

[11] 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.

[12] 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.

telephone interview. Respondents will have some familiarity with the study, which should increase the likelihood of participation. The subcontractor has also achieved similar response rates on comparable follow-up surveys. In addition, a variety of proven effective approaches will be used to maximize response and minimize nonresponse bias.

One strategy for increasing response rates requires that telephone interviewers selected for the project demonstrate a combination of interviewing experience and high-level training focused on encouraging participation. Project-specific training will address the study's purpose and goal, the data collection instrument, and best practices in data collection, while reinforcing concepts for eliminating bias and remaining sensitive to at-risk and special populations. Experienced supervisors will closely monitor all interviewers periodically throughout data collection.

In the sections that follow, we describe studies where the subcontractor has achieved comparable response rates, and we discuss other efforts that we will undertake in order to increase respondent participation.

1.   Comparable Follow-Up Survey Response Rates Achieved by Subcontractor

Achieving high response rates to follow-up surveys can require innovative approaches to data collection, as well as experience tracking respondents over time. The three examples below highlight Mathematica's ability to successfully attain high response rates in follow-up surveys. In the U.S. Department of Labor's Evaluation of Individual Training Account (ITA) Demonstration, Mathematica conducted a follow-up survey of 4,800 randomly selected WIA customers. The survey was conducted about 15 months after random assignment. The questionnaire was administered via computer-assisted telephone interviewing (CATI). To increase response rates, cases in which the sample member could not be interviewed on the telephone were sent to local field interviewers for follow-up. After cooperation was obtained, field interviewers called into the Survey Operations Center (SOC) on cell phones, and the ITA customer was interviewed via CATI. Mathematica achieved an 82 percent response rate to the survey, without the use of an incentive.

Mathematica was also able to attain similarly high response rates over time in other studies. In the Social Security Administration's Services to Evaluate Youth Transition Demonstration Projects, Mathematica developed and evaluated youth transition demonstration (YTD) projects, which are intended to help young people with disabilities make the transition from school to work. YTD projects have been fully implemented in 10 sites across the country. In total, 5,273 youth are participating in the evaluation. The

one-year follow-up survey has been completed in the three original evaluation sites with a response rate of 89 percent.

In addition to the projects described above, Mathematica has also been successful in achieving high response rates in a study with follow-up periods similar to the Green Jobs and Health Care Impact Evaluation. In the Rural Welfare-to-Work Strategies Demonstration Evaluation, two rounds of follow-up surveys were conducted - at 18 and 30 months after random assignment. Mathematica achieved response rates as high as 87 percent in the 18 month survey. In the 30 month survey, interviews were attempted with all sample members, whether or not they had completed an 18-month interview. For the 30-month interview, Mathematica achieved response rates over 82 percent.

In order to achieve high response rates in longitudinal studies, it is critical to complete interviews with respondents who fail to complete an interview in a prior wave. There are a multitude of reasons for nonresponse, ranging from respondents refusing to participate to an inability to locate the respondent. Mathematica prides itself on its ability to overcome these challenges as exemplified in the Evaluation of the Trade Adjustment Assistance Program project where Mathematica successfully conducted follow-up interviews with nearly a third of those who had failed to be interviewed during the initial interview period 15 to 33 months prior.

2. Strategic Use of Technology Tools

Strategic use of technology tools, including CATI, an automated call scheduler, and a survey management system will help to maximize contact with study participants. CATI allows interviewers to move swiftly through the survey instrument, asking only those questions that are relevant to a particular respondent, based on his or her earlier answers. This reduces the length of time respondents spend on the phone, and minimizes the likelihood of respondent fatigue playing into incomplete item responses or surveys.

Language accommodations. To minimize nonresponse due to language barriers, there will be Spanish-speaking interviewers available to assist in conducting the interviews, as needed.

Use of respondent payments. Advance letters signed by a DOL official will be sent by first class postal mail to study participants to convey the importance of the 18-month telephone survey and participation. The letters will include information about a $25 payment being offered for completing this survey. This effort will be repeated for the 36-month follow-up interview. The payment encourages sample members to participate in the survey and provide updated contact information, especially during the 18 months between the first and second follow-up surveys.

Extensive locating efforts. Missing contact information will be obtained either from a private vendor (such as Accurint, a division of LexisNexis) or by the evaluation team's in-house locating staff, who are highly skilled in searching specialized online databases. When necessary, field locating will also be conducted by trained locators. To facilitate efficient contact during the 36-month follow-up interview, a series of questions to obtain additional contact information will be included at the end of the 18-month telephone interview. In addition, study staff will use reminder postcards to obtain updated addresses for sample members who have moved since the 18-month interview.

Nonresponse analysis. The analysis will include the construction of weights for each treatment and control group in each site to account for non-response, as described above in the B.2 Estimation Procedures section, in an effort to minimize the risk of nonresponse bias. If the response rate among the eligible sample is lower than the 80 percent that is expected, the nonresponse analysis will include an estimate of potential bias and the extent to which weights correct for the potential bias.

3. Data Reliability

The two telephone surveys are unique to the current evaluation and will be used with all study participants. Using the same data collection instrument across all study participants will ensure consistency in the collected data and no artificial differences arise between treatment and control group responses that would distort our measures of program impacts. The two telephone surveys will have been reviewed extensively by evaluation team staff and staff at ETA and will be thoroughly tested in a pretest involving nine or fewer individuals from nonparticipating sites. Telephone interviewing staff will complete training on each item in both of the surveys to ensure staff understand each item and record the information accurately.

D. Tests of Procedures or Methods

To assess the clarity of content and wording of the surveys, respondents' burden time, and potential sources of measurement error, study staff conducted a small pretest of the follow-up survey instrument between June 22 and June 28, 2012. A total of 9 pre-test respondents recruited from One-Stop Centers and community training organizations in four metropolitan areas (Washington, DC; Oakland, CA; New Brunswick, NJ; and Chicago, IL). Each pretest respondent was given $25 for participating in the survey. After each interview was conducted, project staff debriefed the participant using a standard debriefing protocol to determine if any words or questions were difficult to understand and answer. No major problems were uncovered in the pre-test. However, some minor formatting and wording changes were made as a result of the test. A memo detailing the pilot test results is included as Attachment 4.

E. Individuals Consulted on Statistical Methods

Consultations on the statistical methods used in this study have been undertaken to ensure the technical soundness of the research. The following people were consulted in preparing this submission to OMB:

As noted earlier, consultations on the research design, sample design, and data collection procedures were part of the study design phase of the evaluation. The purposes of these consultations were to ensure the technical soundness of the study and the relevance of its findings and to verify the importance, relevance, and accessibility of the information sought in the study.

Peer Review Panel Members

- Ms. Maureen Conway, Executive Director, Economic Opportunities Program, Aspen Institute

- Dr. Harry J. Holzer, Professor, Georgetown Public Policy Institute

- Dr. Robert J. LaLonde, Professor, The Harris School, University of Chicago

- Mr. Larry Orr, Larry Orr Consulting

- Dr. Burt S. Barnow, Amsterdam Professor of Public Service, The Trachtenberg School of Public Policy and Public Administration, George Washington University

- Ms. Mindy Feldbaum, Director for Workforce Development Programs, National Institute for Work and Learning

F. Contact Information for Abt Associates

Abt Associates Inc., with its partner Mathematica Policy Research, Inc, is conducting the impact evaluation of the Pathways Out of Poverty and Health Care and Other High Growth Training grants. Contact information for the project director at Abt Associates is as follows:

Dr. Stephen Bell
4550 Montgomery Avenue, Suite 800N
Bethesda, MD 20815
(301) 634-1721
Stephen_Bell@abtassoc.com

REFERENCES

Abt Associates, Inc. "Green Jobs and Health Care Impact Evaluation: Evaluation Design Report." Submitted to the U.S. Department of Labor. Bethesda, MD: Abt Associates, Inc., December 2011.

Bureau of Labor Statistics. "An Occupational Analysis of Industries with Employment Gains: Occupational Employment Statistics (OES) Highlights." Washington, D.C.: BLS, 2010.

Bloom, Howard S., Larry L. Orr, George Cave, Stephen H. Bell, and Fred Doolittle. "The National JTPA Study: Title II-A Impacts on Earnings and Employment at 18 Months." Bethesda, MD: Abt Associates, Inc., 1993.

Heinrich, Carolyn J., Peter R. Mueser, and Kenneth R. Troske, "Workforce Investment Act Non-Experimental Net Impact Evaluation," Final Report, Employment and Training Administration, U.S. Department of Labor, March 2009 (ETAOP 2009-10).

Maguire, Sheila, Joshua Freely, Carol Clymer, Maureen Conway, and Deena Schwartz. "Tuning Into Local Labor Markets: Findings from the Sectoral Employment Impact Study." Philadelphia, PA: Public/Private Ventures, 2010.

Mills, Gregory, Daniel Gubits, Larry Orr, David Long, Judith Feins, Bulbul Kaul, Michelle Wood, Amy Jones and Associates, Cloudburst Consulting, and the QED Group. "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., 2006.

Schochet, Peter Z. "An Approach for Addressing the Multiple Testing Problem in Social Policy Impact Evaluations." *Education Review,* vol. 33, no. 6, pp. 9-149, 2009.