

## **PART B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS**

In this document, we discuss the statistical methods to be used in the data collection activities for the *Evaluation of Strategies Used in the TechHire and Strengthening Working Families Initiative (SWFI) Grant Programs*. This study is sponsored by the Chief Evaluation Office (CEO) and Employment and Training Administration (ETA) in the U.S. Department of Labor. The purpose of the evaluation is to identify whether the grants help low-wage workers obtain employment in and advance in H-1B industries and occupations and, if so, which strategies are most helpful. CEO has contracted with Westat and its subcontractor, MDRC, to conduct this evaluation. The evaluation includes three components: an implementation study, a randomized controlled trial (RCT) study, and a quasi-experimental design (QED) study.

This request for clearance includes the following data collection activities:

- Baseline Information Form (BIF)
- 6-month follow-up survey
- Site visit interviews with grantee staff
- Site visit interviews with grantee partners
- Participant tracking form

Subsequent OMB submissions will seek clearance of additional data collection activities, including a grantee survey, grantee semi-structure interviews, a partner information template, a partner survey, partner semi-structure telephone interviews, a second round of site visits, and an 18-month follow-up participant survey.

### **B.1 Respondent Universe and Sampling Methods**

#### **B.1.1 Respondent Universe**

##### **Baseline Information Form**

DOL is selecting the six grantees to participate in the RCT based on several criteria including their ability to generate needed sample sizes, innovative training models, capacity, and other factors. Sites will use their existing eligibility criteria to identify people who are eligible for the program. The respondent universe consists of all individuals who are randomized to participate in the TechHire or SWFI treatment or the control group during the study intake period (between January 2018 and December 2019). To be randomly assigned for the chance of participating in TechHire or SWFI, an individual must apply to and be accepted at one of the 6 TechHire or SWFI programs and provide consent to participate in the study. Therefore, the respondent universe is the subset of eligible individuals who apply to the 6 TechHire or SWFI programs and agree to participate in the study. These individuals will be asked to complete the Baseline Information Form (BIF) and Informed Consent Form (ICF).

Each RCT site is expected to enroll 600 individuals, 300 in the treatment group and 300 in the control group, for a total of 3,600 participants overall (see table B.1.1).

### 6-Month Follow-Up Survey

The universe of respondents for the 6-month follow up survey is all individuals who are in the treatment and control groups. We do not plan to select a sample.

We expect an 80 percent response rate to both the 6-month follow-up survey.

**Table B.1.1 Expected Number of Treatment and Control Group Members and Survey Respondents**

		<b>Treatment</b>	<b>Control</b>	<b>Total</b>
Expected Number of Respondents to Baseline Information Form	Site 1	300	300	600
	Site 2	300	300	600
	Site 3	300	300	600
	Site 4	300	300	600
	Site 5	300	300	600
	Site 6	300	300	600
	All Sites	1,800	1,800	3,600
Expect Number of Respondents to 6 - Month Survey	Site 1	240	240	480
	Site 2	240	240	480
	Site 3	240	240	480
	Site 4	240	240	480
	Site 5	240	240	480
	Site 6	240	240	480
	All Sites	1,440	1,440	2,880

### Site Visits

The study team will conduct sites visits to the six RCT sites. Each visit will be 2-3 days in length, and will include hour interviews with each staff and observations of program operations. The study team will interview grantee program staff, such as the Program Director, Program Manager/Coordinator, Data Analysts, and other support staff at each grantee site. The study team will also interview grantee partners, including training partners and employer partners. Interviews will be conducted with 10 grantee program staff and 8 partners during each site visit, for a total of 60 grantee staff and 48 partners. Each interview will take one hour. We will not use any statistical methods in the selection of grantee staff or partners.

### Participant Tracking Form

The universe of respondents for the 6-month follow up survey is all individuals who are in the treatment and control groups. We do not plan to select a sample.

## **B.2 Procedures for the Collection of Information**

### **B.2.1 Statistical Methodology for Stratification and Sample Selection**

Of the 53 grantees, 6 will be part of the RCT study. Grantees are being selected purposively. This selection process considers three general criteria: (1) the strength of the proposed program intervention and the demonstrated capability of the grantee to carry it out; (2) the readiness of the grantee to implement the research procedures, including the grantee's ability to reach specific sample size goals, to maintain a clear contrast between the program and control groups, and to maintain a strong adherence to collecting high-quality data; and (3) cross-site research considerations relating to the mix of strategies, population targets, implementing institutions, and industry and occupational targets. Some of these cross-site considerations include:

- The relative importance of other features such as grantees that are undertaking different programs so that the RCTs cover a spectrum of, for example, training approaches and populations, have a range of locations within a grantee, use different types of providers, or are located in different regions of the country; and
- The value of identifying grantee sites with similar interventions (e.g., coding boot camps or work-based learning) to pool the samples across sites, in the event sites are not able to meet agreed-upon sample targets.

All participants who are eligible, complete the BIF, and consent will be randomly assigned to a treatment or control group. The study intake period will be from January 2018 to December 2019. We anticipate that 3,600 individuals will be randomly assigned. All individuals will be included in the follow-up data collection effort.

### **B.2.2 Estimation Procedures**

#### **Baseline Information Form**

Analysis of data from the BIF will be purely descriptive and designed to document characteristics of participants in the programs. Because there is no sampling, the data analysis will not be estimates.

#### **6-Month Follow-Up Survey**

The objective of the RCT 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 six RCT study sites. Specifically, the RCT will identify the extent to which training and supportive services in each site improve participant employment and earnings.

This section describes our analytic approach to estimate program impacts using the 6-month survey data that are included in this request for clearance, as well as the National Directory of New Hires (NDHD) data. The 6-month survey is intended to capture impacts on intermediate

outcomes of educational and training completion and removal of work-related barriers, including transportation and childcare. Key outcome variables include training completion, industry of training, credential attainment, has regular child care arrangement, receives child care subsidy, ever worked, currently working, and current/most recent job industry, hours worked, and hourly wage.

Given that random assignment produces balance in expectation, the basic impact estimates can be computed using the difference in mean outcomes between the treatment group members and control group members in each of the 6 RCT sites. Each of the resulting estimates for each outcome is unbiased because the individuals who comprise the treatment and control groups in the site were assigned at random to either the treatment or control from a common pool and hence are expected to be statistically equivalent on all factors at baseline. As a result, any statistically significant differences in outcomes between the groups can be attributed to the effects of the intervention. Regression analysis that controls for background characteristics, including prior earnings and employment, education, and demographics, will be used to improve the precision of the impact estimates. A standard linear regression will be used to estimate impacts. This approach is used in all impact analyses that we are aware of over the past 40 years. The rationale for this standard approach, of using inferential (rather than descriptive) statistics, rests on the goal to generalize the results to a super population of TechHire and SWFI grantees. As discussed in the site selection section, even though probability sampling will not be used to select the sites, an effort was made to ensure that the grantees chosen for the impact analysis cover a range of program features, geographic regions, and populations. The Department of Labor hopes to learn not only about these specific six grantees, but to generalize these findings to the dozens of TechHire grantees throughout the nation.<sup>1</sup>

A key question to address is whether impact estimates will be at the site level, or whether sites will be pooled. Some of the grantees may not be able to obtain adequate sample sizes within the currently planned intake period. If viable and necessary (due to small sample sizes), the data will be pooled across sites. The extent of pooling across programs/grantees will depend on their equivalency. While it would be ideal to have site specific estimates, pooling sites may be necessary. In fact, in many studies that show site-specific impact estimates, the pooled estimate is typically included because it provides a bottom-line indicator across a range of sites. Depending on the policy question, the pooled estimate can be quite useful. Ultimately, DOL will make the decision to show site specific or pooled estimates (or both) after assessing the RCT grantees' available sample sizes, target populations, and programmatic approaches but before we estimate impacts. A middle option is to cluster some of the sites—for example, to group by

---

<sup>1</sup> Bloom et al, 2017 assert that these types of inferences to a super population are appropriate even in the case of convenience sampling: “Even in a convenience sample, sites are usually chosen not because they comprise a population of interest but rather because they represent a broader population of sites that might have participated in the study or might consider adopting the program being tested. Hence, the ultimate goal of such studies is usually to generalize findings beyond the sites observed, even though the target of generalization is not well-defined.”

TechHire and SWFI grantees. Finally, we note that it is very common for evaluations to “lead” with a pooled estimate as the primary estimator and then show impacts by sites for sites that can stand on their own. Such an approach was used in the P/PV Sectoral Employment Impact Study (McGuire et al, 2010). If we do end up estimating impacts for the pooled sample, we will add site indicator dummies to the linear regression models in order to account for site fixed effects.

Impacts will be calculated for key subgroups to better understand what works best for whom. In impact studies, subgroup impacts have been estimated several different ways. In “split-sample” subgroup analyses, the full sample is divided into two or more mutually exclusive and exhaustive groups (for example, by gender or for those with more severe barriers and less work experience). In this approach, impacts are estimated for each group separately. In addition to determining whether the intervention had statistically significant effects for each subgroup, Q-statistics are used to determine whether impacts differ significantly across subgroups (Hedges & Olkin, 1985). Regardless of the exact subgroup estimation strategy, it is clear that subgroups would have to be computed on the pooled sample rather than with sites.

We will strive to limit subgroup comparisons to those for which theory and prior studies provide good reasons for expecting subgroup differences on advancement outcomes. This is to guard against the chance of a “false positive,” which stems from the fact that the more subgroups that are examined, the greater the chance of finding one with a large effect, even when there are no real differences in impacts across subgroups. We will also consider the use of corrections for multiple comparisons, such as the Bonferroni or Benjamini-Hochberg corrections.

In addition to traditional subgroups based on background characteristics, nonexperimental subgroup approaches such as examining the effects of different treatment elements will be explored. Methods such as instrumental variables or regression-based subgroups as done by Peck (2013) or Kemple and Snipes (2000) will be used, but this would be a secondary, exploratory analysis given that the effect estimates will be less reliable.

### **Site Visits**

Data collected from the site visits will provide in-depth qualitative information about implementation at the 6 RCT sites; no estimate procedures will be used. Analysis will be purely descriptive.

### **Participant Tracking Form**

The participant tracking form will be used to obtain updated contact information for participants. There will be no analysis or estimation.

### B.2.3 Degree of Accuracy Needed

Based on the selection of 6 grantees for the RCT, we assume a total of 600 participants (300 treatment and 300 control) for estimating burden. Table B.2.1 shows the Minimum Detectable Effects (MDEs) for a few different sample sizes:<sup>2</sup>

- 1) A grantee level sample size of 600 total
- 2) A grantee level sample size of 400 total
- 3) A scenario in which we pool 2 sites with 400 total each

In addition to the sample size of 600 total for burden purposes, we calculated MDEs for a sample size of 400 total in the event that one or more grantees are unable to meet the sample size target of 600. Within each of these three sample sizes, MDEs are shown based on the expected sample sizes for both the administrative records data (which are assumed to cover all sample members) and the survey data (which are assumed to cover 80 percent of sample members).<sup>3</sup> The four rightmost columns of the table show the MDEs for impacts on percentage measures, such as employment, assuming two different standard deviations (0.4 and 0.5)<sup>4</sup> and continuous measures, such as earnings, assuming two different standard deviations (\$2,000 and \$3,500).<sup>5</sup>

---

<sup>2</sup> The minimal detectable effects were estimated using the following formula from Bloom et al. 2006:

$$MDE(\bar{Y}_T - \bar{Y}_C) = M_{n-2} \sqrt{\frac{\sigma^2}{nP(1-P)}}$$

Here the Y terms refer to the means for treatment and control group members (subscripted by T and C). M is the multiplier, which is a constant that factors in the statistical power (here 0.8), the statistical significance (here 0.1), and the number of degrees of freedom in the impact analysis equation. In the two tailed test scenario that we have used here,  $M = 2.49$ . The term on the right is the standard error which is computed by dividing the variance ( $\sigma^2$ ) by the random assignment ratio (P).

<sup>3</sup> The MDESs presented are for a two-tailed test at 0.10 significance level with 80 percent power. These MDES assume 50 percent of the sample is assigned to the treatment group and 50 percent is assigned to the control group, that 10 covariates will be used, and that the covariates will have a weak relationship with the outcomes (R-squared=0.15). This R-squared value is somewhat low compared to what's been observed in other training and employment studies, and is intentionally a fairly conservative estimate. If the covariates actually explain more of the variation in outcomes, this will only increase the power of the study.

<sup>4</sup>A standard deviation of 0.5 assumes the worst case scenario. The point of maximum variance for a percentage measure is 0.5 (a control group level of 50 percent). At that point, an MDES of 0.2 translates into an MDE of 10 percentage points. The further the variance is from 0.5, the smaller the MDE. For example, if the control group level for a measure is 20 percent, the MDE for a study powered at 80 percent would be 8 percentage points.

<sup>5</sup>These standard deviations are based on standard deviations from MDRC's evaluation of the WorkAdvance program (Hendra et al., 2016).

**Table B.2.1. Minimum Detectable Effects for Key Outcomes, by Sample**

Sample	Sample Size	MDE <sup>a</sup>	MDE <sup>a</sup>			
			Employment		Annual Earnings	
			SD = 0.4	SD = 0.5	SD = 8,000	SD = 14,000
Assuming 600 sample members per site						
Administrative records	600	0.187	7.5	9.4	1,496	2,618
Survey	480	0.210	8.4	10.5	1,680	2,940
Assuming 400 sample members per site						
Administrative records	400	0.230	9.2	11.5	1,840	3,220
Survey	320	0.257	10.3	12.9	2,056	3,598
Assuming pooling 2 sites with 400 sample members per site						
Administrative records	800	0.162	6.5	8.1	1,296	2,268
Survey	640	0.181	7.2	9.1	1,448	2,534
Assuming pooling 2 sites with 600 sample members per site						
Administrative records	1,200	0.132	5.3	6.6	1,056	1,848
Survey	960	0.148	5.9	7.4	1,184	2,072

<sup>a</sup> Minimum Detectable Effect

MDES for grantee's with a sample size of 600 total are acceptable for inclusion in a site specific analysis. As discussed above, a common threshold for a well powered estimate is that MDEs should be able to detect impacts which are smaller than 0.210 standard deviations different from the control group level. At a sample of 600, impacts measured from administrative records (such as earnings or employment measures) would have to be at least 0.187 standard deviations which is within the threshold. Given the intensity of some of the proposed treatments, the 600 sample size is acceptable for site based analysis.

With a sample size of 400 and 320 for the survey (assuming an 80 percent response rate), the MDESs are 0.23 and .257, respectively. Assuming 50 percent of the control group was employed (that is, the standard deviation is 0.5), these MDESs translate into MDEs of between 11.5 and 12.9 for percentage measures. MDES for earnings measures are also shown. Depending on the control group levels, earnings impacts would have to be in the \$2,000-\$3,000 range to have a reasonable chance of being statistically significant. These MDEs are *quite high* and based on past studies it would be unreasonable (though certainly not impossible) to expect impacts this large except for a few outlier sites. For this reason, if sites have sample sizes in the 400 range they'd have to be pooled with other sites in order to be included in the RCT analysis of earnings.

Table B.2.1 also shows sample sizes for two pooled samples. Under the first scenario we would pool 2 sites with four hundred sample members each. The resulting analysis would be powered to detect impacts on employment in the 8 or 9 percentage point or \$1500-\$2000 range.

Power would be more than adequate in a scenario in which we pool 3 sites of 400 each (or 2 sites of 600 each). Under that scenario MDEs are in the more standard range of 0.132 (for administrative records) which translates into employment effects in the 6 or 7 point range and earnings effects between \$1000-\$2000.

#### **B.2.4 Unusual Problems Requiring Specialized Sampling Procedures**

None.

#### **B.2.5 Any use of periodic (less frequent than annual) data collection cycles to reduce burden.**

There is no use of less than annual data collection cycles because of the relatively short duration of the grant programs being evaluated.

### **B.3 Methods to Maximize Response Rates and to Deal With Issues of Nonresponse**

#### **B.3.1 Methods to Maximize Response Rates**

##### **Baseline Information and Consent Forms**

Baseline information and informed consent will be collected from all study participants at the six RCT sites following determination of program eligibility, but prior to random assignment. These data will provide general information about participants that will ensure the comparability of the program and control groups, obtain data needed for subgroup analyses, and facilitate contact for subsequent follow-up surveys. Baseline data (including contact information) and informed consent will be collected via MDRC's web-based random assignment system.

The Informed Consent Form will be used in the six RCT sites. It will ensure that participants: 1) understand the TechHire/SWFI evaluation, as well as their role and rights within the study; and 2) provide their consent to participate. As mentioned, informed consent will be collected via MDRC's online random assignment system. Individuals will check a box indicating whether they consent to participate in the evaluation. This will serve as an "electronic signature."

The Informed Consent Form has a eighth-grade readability level and draws on consent forms currently in use in other studies. It meets all human subjects requirements. To ensure that all study participants receive a clear, consistent explanation of the project, the evaluation team will provide program staff with a script and talking points to summarize and address questions that may come up from reading the consent form. Staff will be trained on how to introduce and discuss the goals and design of the project, the random assignment process and data collection



efforts; all staff will emphasize: 1) that participation in the study is voluntary; 2) the benefits and risks of participating in the study, and 3) that strict rules are in place to protect sample members' privacy. The informed consent form indicates that participants cannot access TechHire or SWFI services if they do not consent to participate in the study or if they withdraw from the study.

The baseline form will collect basic identifying information about sample members in the six RCT sites, including social security number and date of birth. It also includes demographic items such as race and ethnicity, language spoken, and housing and marital statuses. Finally, the form also includes questions about participants' education and employment history, use of public assistance, need for child care services, and any involvement with the criminal justice system. These data will be used by the evaluation team to ensure the comparability of program and control groups, to facilitate contact for follow-up surveys, and to inform subgroup analyses in the impact study.

As noted earlier, baseline data will be collected via MDRC's web-based random assignment system. To help ensure quality responses, the system will be set up to impose certain rules for accepted values, control skips, and prompt users with various messages if they report potential errors or do not answer all of the questions. As an example, the system will require that a key identifier—Social Security number—be entered twice, and it will return an error message that must be addressed before proceeding if the entries are not identical or do not meet The Social Security Administration's validation rules. Site staff will also be trained on how to offer guidance to individuals on how they should answer some key baseline questions.

Contact information will also be collected prior to random assignment. In addition to their own name, address, and phone numbers, participants will be asked to provide information about three additional individuals who are likely to be in future contact with them and can assist the research team to locate them for follow-up surveys. While individuals are not required to report this information, staff will be instructed on how to explain the importance of collecting this information to individuals.

We expect to achieve a 100 percent response rate to the BIF. Westat has achieved a 100 percent or higher response rates for similar studies. For example, Westat's Mental Health Treatment Study (MHTS), a randomized controlled trial of supported employment services for people with serious mental illness funded by Social Security Administration (SSA), achieved a 100 percent response rate to the BIF.

Missing values for baseline variables that are used as covariates will be imputed using the full sample's mean. Dummy variables will also be added to the regression models to indicate missing status for each of the covariates with missing values.

Sample members assigned to the control group will receive \$50.00 as a token of appreciation for the comprehensive intake and screening process that could require several return visits to the grantee's offices. This will not increase the response rate to the BIF because the universe for the BIF is individuals who are randomly assigned.

### **6-Month Follow-Up Survey**

As discussed in Part A, we propose to implement an “early bird” incentive model for the 6-month follow-up survey. Participants will be offered \$30 for completing the survey within the first four weeks and \$20 for completing it after that time.

The sequential multi-mode data collection approach includes the administration of a web survey to study participants (treatment and control groups) in the six RCT sites followed by telephone calls to nonrespondents to complete a computer-assisted telephone interview (CATI). Our goal for the 6-month follow-up survey is an 80 percent response rate. We are assuming that we will obtain 25 percent of these completes from the web survey and the remaining 75 percent from the CATI data collection. Alternatively, the 80 percent response rate is composed of 20% by web and 60% by CATI.

Westat has achieved 80 percent or higher response rates for similar studies. For example, the MHTS included seven quarterly follow-up surveys with RCT study participants in the treatment and control groups after random assignment. Westat achieved over 80 percent response rates in all seven quarters, including over 80 percent response rates for both the treatment and control groups for each quarterly survey prior to quarter 6.

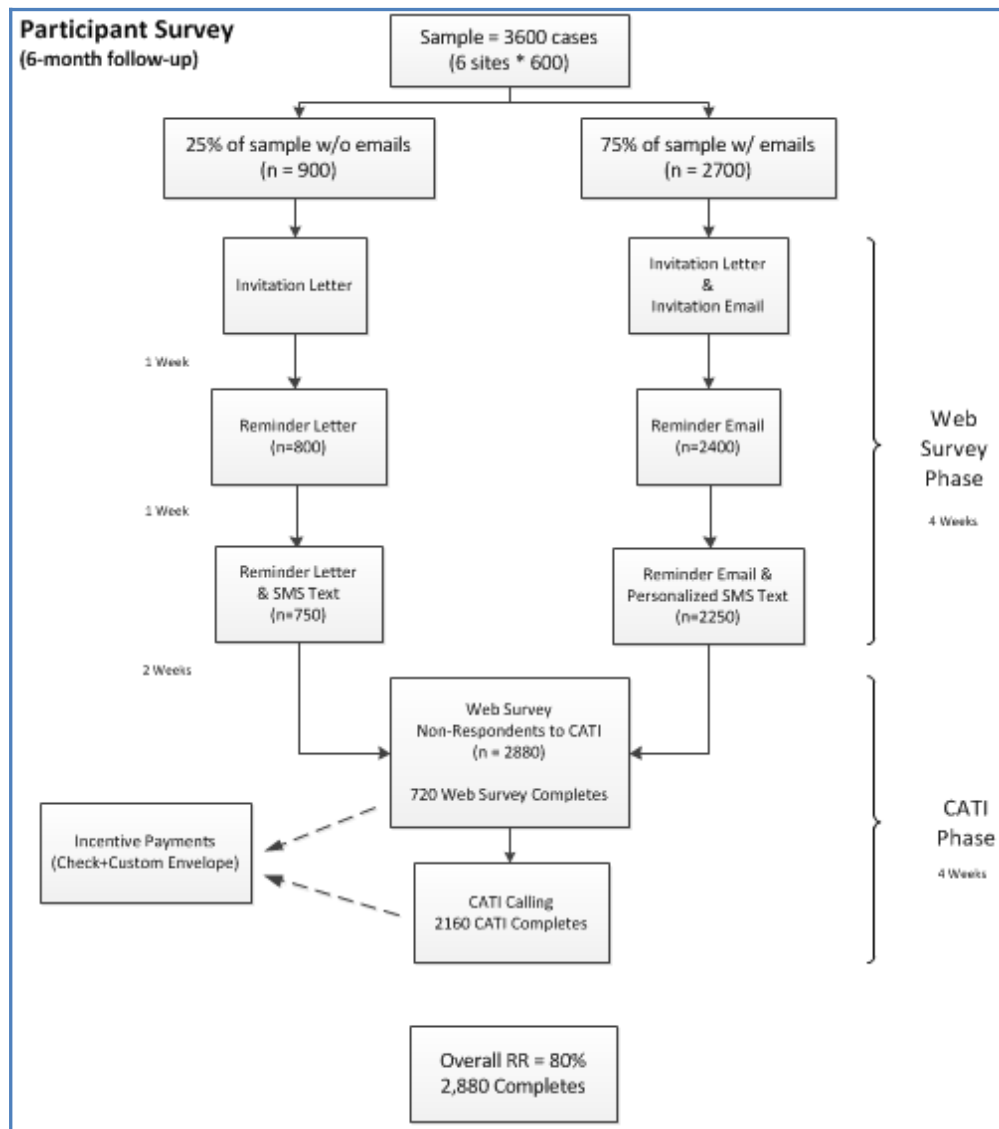
Figure B.3.1 provides a flowchart of the data collection methodology to be used for both the 6-month follow-up survey. Cohorts will be released for survey fielding each month. Participants will be mailed an invitation letter via first class postal mail with study information that contains the web survey URL and a unique PIN to access the survey. A set of frequently asked questions (FAQs) will also be included on the back of the invitation letter along with a sponsorship letter from DOL to help add credibility to the study, and to encourage participation. The invitation will also provide a phone number and email address for participants to contact Westat for technical assistance. We will also send an invitation email to those cases where we have their email address from the baseline intake phase. Our experience shows that response rates are slightly higher on web surveys when sampled members can click on a personalized web survey link within an email compared to entering a URL from a postal letter. In order to decrease respondent burden and to improve response rates, the web survey will be accessible across a wide array of computer configurations and mobile devices. We are assuming, based on our experience with prior studies, that 75% of our participant survey sample will have email addresses.

One week following the initial invitation mailing, participants will receive the first of three weekly reminders. Participants with email addresses will receive an email reminder, and those without an email address will receive the reminder letter via postal mail encouraging their participation in the survey. The second and final weekly reminder will be sent out to nonrespondents similarly to the first weekly reminder. The final reminder will encourage sample members to complete the web survey and inform them that we will contact those who have not completed the web survey by telephone. Additionally, SMS text messages will also be sent at the same time as the weekly email reminders to those participants who provide permission to text them. Cell phone number and permission to text will be asked in the BIF and 6-month follow-up survey.

### **Participant Tracking Form**

Participants will be provided with a pre-paid \$2 incentive for completing and returning the contact information postcard (or emailing or calling with the information) to the evaluation team.

**Figure B.3.1 Proposed data collection flow for the 6-Month Follow-Up Survey**



Beginning the 5<sup>th</sup> week of the data collection field period for each monthly cohort, nonrespondents to the web survey will move on to the telephone phase of the study where participants will be contacted using our CATI method to complete the participant survey. For efficiency purposes and to minimize any mode effects, the web survey will undergo minimal revision for telephone administration by trained interviewers. The telephone instrument version of the survey will be identical to the web version, with some format adaptations appropriate for interviewer administration such as transitions between survey sections when content changes. The web survey will remain open to respondents during the CATI phase, although no additional reminders will be sent out. We expect the CATI interviews to be in the field approximately 4 weeks for each monthly cohort in order for all assigned cases to get through the CATI calling algorithm and for us to reach our completion goal.

We expect to get about 2,160 completed CATI surveys, yielding approximately 2,880 total completed surveys (web + CATI). The data management system for this multi-mode study will include automated systems for managing and tracking respondent information and survey completions.

Westat will use established procedures for tracing study participants. All postal non-deliverables (PNDs) received from the survey mailings will receive individual-level tracing using LexisNexis and the National Change of Address database maintained by the U.S. Postal Service. Contact information for three people who know the respondent well, which will be collected in the BIF and 6-month follow-up survey, will be used to trace participants. Finally, if participants cannot be located via telephone during the 4-week CATI period, field data collectors will be deployed to the last known address and use field tracing techniques to attempt to locate the participant.. If the participant can be located, field locators will administer the survey in-person.

### **B.3.2 Nonresponse Bias Analysis**

Even though intensive methods will be used to increase response rates and convert non-responders to the survey, non-response bias is still a concern. Survey nonresponse can bias the impact estimates if the outcomes of survey respondents and nonrespondents differ, or if the types of individuals who respond to the survey differ across the program and control groups. We will use several methods to assess the effects of survey nonresponse during data collection and using data collected for the study.

During data collection, we will take steps to understand, monitor, manage and address potential sources of non-response bias. During the survey fielding period, we will review contact attempts and disposition status which will enable us to monitor response rates by cohort (defined by time of random assignment), research group, and site. Should significant gaps in response rates among these groups occur, we will intensify recruitment efforts for the affected group. These intensified efforts will include prioritizing the efforts of the most experienced survey interviewers towards the affected group.

We will also examine nonresponse using data collected for the study. First, we will use baseline data (which will be available for the *full* research sample) to conduct statistical tests (chi-squared and t-tests) to gauge whether treatments who respond to the interviews are fully representative of all treatment group members, and similarly for control group members. Noticeable differences in the characteristics of survey respondents and nonrespondents could suggest the presence of nonresponse bias. Furthermore, we will test whether the baseline characteristics of *respondents* in the two research groups differ from each other. Although baseline characteristics for the full sample should not differ much between the program and control groups, significant differences between program and control group respondents could mean that impacts estimated from surveys will confound program impacts with pre-existing differences between the groups.

Second, we will assess nonresponse bias using administrative records data. For example, we will examine whether *impacts* on employment rates differ for survey respondents and survey nonrespondents. If program impacts are substantially different for respondents and nonrespondents, that would make us more cautious about drawing conclusions from the survey.

We will use several approaches to correct for potential nonresponse bias in the estimation of program impacts. First, we will adjust for observed differences between treatment and control group respondents using regression models. Second, because this regression procedure will not correct for differences between respondents and nonrespondents in each research group, we will construct sample weights so that the weighted observable baseline characteristics of respondents are similar to the baseline characteristics of the full sample of respondents and nonrespondents. We will construct weights for treatment and control group members using the following three steps:

1. Estimate a logit model predicting interview response. The binary variable indicating whether or not a sample member is a respondent to the instrument will be regressed on baseline measures.
2. Calculate a propensity score for each individual in the full sample. This score is the predicted probability that a sample member is a respondent, and will be constructed using the parameter estimates from the logit regression model and the person's baseline characteristics. Individuals with large propensity scores are likely to be respondents, whereas those with small propensity scores are likely to be nonrespondents.
3. Construct nonresponse weights using the propensity scores. Individuals will be ranked by the size of their propensity scores, and divided into several groups of equal size. The weight for a sample member will be inversely proportional to the mean propensity score of the group to which the person is assigned.

This propensity score procedure will yield large weights for those with characteristics that are associated with low response rates (that is, for those with small propensity scores). Similarly, the procedure will yield small weights for those with characteristics that are associated with high response rates. Thus, the weighted characteristics of respondents should be similar, on average, to the characteristics of the entire research sample.

It is important to note that the use of weights and regression models adjusts only for *observable* differences between survey respondents and nonrespondents in the two research groups. The procedure does not adjust for potential unobservable differences between the groups. Thus, our procedures will only partially adjust for potential nonresponse bias. We will use administrative data to assess whether such bias is present in our data, as discussed above.

Records with missing values on dependent (or outcome) variables will be excluded from the impact estimates for those variables. This includes records for participants who do not respond to the entire survey, as well as records for participants who do not answer individual questions. As a sensitivity check, we will use multiple imputation of survey outcomes as suggested by Puma et al, 2009.

#### **B.4 Test of Procedures or Methods to be Undertaken**

The 6-month follow-up survey was pretested in April 2017, before submission to OMB. For the pretest, 7 individuals were recruited from the TechHire program at Montgomery College in Rockville, MD and a One-Stop Center in New York City. Pretests were conducted over the phone. Individuals were paid \$50 for their participation in the pretest. The pretest used retrospective probing in which respondents completed the full survey and were then probed about the clarity of the questions and any potential problems with the instrument. The benefit of retrospective probing is that it allowed us to estimate the burden. The version of the instrument submitted to OMB incorporates the pretest results and no changes are anticipated. In terms of burden, the average time to complete the survey was 18 minutes, although the time varied based on whether the respondent was employed, participated in job training, or were parents, as these determined major skip patterns. No testing of the site visit protocols will be conducted but they will be refined based on responses to questions obtained during the initial site visit interviews.

#### **B.5 Individuals Consulted on Statistical Methods and Individuals Responsible for Collecting and/or Analyzing the Data**

##### **B.5.1 Individuals Consulted on Statistical Methods**

The following people were consulted on statistical methods in preparing this submission to OMB:

##### ***Westat***

Dr. Joseph Gasper (240) 314-2470  
Dr. Frank Jenkins (301) 279-4502  
Dr. Frank Bennici (301) 738-3608

##### ***MDRC***

Dr. Richard Hendra (212) 340-8623  
Dr. Barbara Goldman (212) 340-8654

We have also assembled a Technical Working Group (TWG) consisting of five experts in the following areas: (1) training low-skill/low-income people; (2) experimental design; and (3)

quasi-experimental design. The TWG will be consulted on the evaluation design, site selection, instruments, and analysis.

Technical Working Group Members

- Kevin M. Hollenbeck, Ph.D., Vice President, Senior Economist, W.E. Upjohn Institute
- Jeffrey Smith, Ph.D., Professor of Economics, University of Michigan
- Gina Adams, Senior Fellow, Center on Labor, Human Services, and Population at The Urban Institute
- David S. Berman, MPA, MPH, Director of Programs and Evaluation for the NYC Center for Economic Opportunity, in the Office of the Mayor
- Mindy Feldbaum, Principal at the Collaboratory

**B.5.2 Individuals Responsible for Collecting/Analyzing the Data**

*Westat*

Dr. Joseph Gasper (240) 314-2470

Mr. Wayne Hintze (301) 517-4022

**MDRC**

Dr. Richard Hendra (212) 340-8623

Ms. Alexandra Pennington (212) 340-8847

Ms. Kelsey Schaberg (212) 340-7581



## References

- Bloom, H. S. (2006). *The Core Analytics of Randomized Experiments for Social Research*. New York, NY: MDRC.
- Bloom, H. S., Raudenbush, S. W., Weiss, M. J., and Porter, K. (2017). Using Multi-Site Experiments to Study Cross-Site Variation in Effects of Program Assignment. *Journal of Research on Educational Effectiveness*, 1-26.
- Dorsett, R., & Robins, P. K. (2013). A multilevel analysis of the impacts of services provided by the UK employment retention and advancement demonstration. *Evaluation review*, 37(2), 63-108.
- Heinrich, C. J., Mueser, P., Troske, K. R., & Benus, J. M. (2008). Workforce Investment Act non-experimental net impact evaluation. *Report to US Department of Labor. Columbia, MD: IMPAQ International*.
- Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press
- Hendra, R., Greenberg, D. H., Hamilton, G., Oppenheim, A., Pennington, A., Schaberg, K., & Tessler, B. L. (2016). Encouraging Evidence on a Sector-Focused Advancement Strategy: Two-Year Impacts from the WorkAdvance Demonstration.
- Hollenbeck, K., & Huang, W. J. (2008). Workforce program performance indicators for the commonwealth of Virginia.
- Kemple, J. J., & Snipes, J. C. (2000). Career Academies: Impacts on Students' Engagement and Performance in High School.
- Maguire, S., Freely, J., Clymer, C., Conway, M., & Schwartz, D. (2010). Tuning in to Local Labor Markets: Findings from the Sectoral Employment Impact Study. *Public/Private Ventures*.
- Peck, L. R. (2013). On analysis of symmetrically predicted endogenous subgroups part one of a method note in three parts. *American Journal of Evaluation*, 34(2), 225-236.
- Puma, M. J., Olsen, B., Bell, S. H., & Price, C. (2009). *What to Do When Data Are Missing in Group Randomized Controlled Trials* (NCEE 2009-0049). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute for Education Sciences, U.S. Department of Education.
- Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American statistical Association*, 79(387), 516-524.
- Stuart, E. A., Huskamp, H. A., Duckworth, K., Simmons, J., Song, Z., Chernew, M. E., & Barry, C. L. (2014). Using propensity scores in difference-in-differences models to estimate the effects of a policy change. *Health Services and Outcomes Research Methodology*, 14(4), 166-182.