Part B: Collection of Information Employing Statistical Methods

The U.S. Department of Labor (DOL) has contracted with Abt Associates and MEF Associates to conduct an evaluation of the Ready to Work (RTW) Partnership Grants (hereafter, RTW Evaluation). DOL is seeking approval from OMB under the Paperwork Reduction Act of 1995 for an 18-month follow-up survey associated with the RTW Evaluation. DOL funded the RTW Grant program in 2014 in order to improve the economic well-being of Americans facing long-term unemployment. It has awarded $169 million to 23 grantees, with individual awards ranging from $3 to $10 million. DOL also sponsored an evaluation, which includes: (1) an implementation study that examines the operation of the programs and participation patterns of program enrollees in key program activities, and (2) an impact study that uses a random assignment research design to determine whether selected grantee programs increased participants' employment, earnings, and other outcomes. In the selected study sites, sample members will be randomly assigned to either a treatment group that will be offered the RTW-funded services, or a control group that cannot receive these services. The four RTW programs selected for the evaluation are:

- Anne Arundel Workforce Development Corporation's (AAWDC) Maryland Tech Connection (MTC) program operating in the Baltimore, Maryland and Washington, DC metropolitan region.
- Jewish Vocational Service's (JVS) Skills to Work in Technology program operating in the San Francisco Bay Area.
- RochesterWorks!'s Finger Lakes Hired program operating in Monroe County, New York, which includes the city of Rochester.
- Worksystems Inc.'s (WSI) Reboot Northwest program operating in the Portland, Oregon and the Vancouver, Washington metropolitan region.

Across the four sites, the evaluation will collect data from a projected 2,510 treatment group members and 2,510 control group members, for total of 5,020 individuals.

OMB approved initial data collection activities for the RTW Evaluation under OMB control number 1205-0507 (original approval on March 1, 2013; two approvals for nonsubstantive changes were approved on July 31, 2013 and March 13, 2015; and an extension with revisions request is currently under review at OMB). Those approved data collection activities included telephone interviews and site visits for the purpose of selecting study sites; study participant consent and Baseline Information Forms (BIF); and implementation study site visit protocols.

This submission seeks OMB approval for an additional data collection instrument, the 18-month follow-up survey of study participants. The survey will provide critical information on the experiences and educational and economic outcomes of study for both treatment and control members. Specific outcomes to be considered include the receipt of training and related supports, receipt of credentials, factors that may affect the ability to work (such as attitudes about work and self and availability of work supports), employment, earnings, job characteristics, receipt of public benefits, and household income. For treatment group members, the survey will also collect opinions about the RTW services provided.

B.1 Respondent Universe and Sampling Methods

The respondent universe for this study consists of a projected 5,020 sample members enrolled in the evaluation from the four RTW grantee programs participating in the evaluation. There were two steps in the selection process of RTW Evaluation participants. First, four grantees were purposively selected from the universe of 23 RTW grantees based on the following criteria:

- ***Strength of the intervention.*** So that the evaluation sample would provide a test of the RTW training model, the evaluator chose grantee programs that contained elements central to the RTW Solicitation for Grant Applications: (1) coaching and other direct job placement services; (2) short-term training that leads to direct job placement; and (3) longer-term career pathway training that culminates in an industry-recognized credential and employment.
- ***Program size and sufficient demand to create a control group.*** Since the evaluation aims to generate site-specific impact estimates, a key site selection criterion was grantees' enrollment goals. In order to detect any policy-relevant differences in outcomes between the treatment and control groups, the evaluation team estimated that grantees must generate samples of approximately 500 individuals for the treatment group within a three-year sample enrollment period, and a similar number for the control group.
- ***Treatment-control contrast.*** In order to isolate the effect of RTW training, the evaluation team assessed whether grantees were in communities with similar training opportunities. There is little chance of detecting impacts if the control group can access training and services similar to those provided by RTW grantees.
- ***Ability to accommodate the study procedures.*** Because grantees included in the evaluation must integrate random assignment into their enrollment procedures, the evaluator selected sites with an intake process conducive to the necessary modifications.

The selected sites were Anne Arundel Workforce Development Corporation (AAWDC); Jewish Vocational Service (JVS); RochesterWorks!; and Worksystems Inc.(WSI)

Second, all eligible and consenting applicants for each of the selected sites' programs will be enrolled in the study sample.[1] Two of the grantees, WSI and JVS, will enroll participants into the study until they reach a total sample of 1,000 participants each (500 treatment participants and 500 control participants at each site). After they reach this goal, they will continue enrolling trainees, but will discontinue random assignment and sample enrollment, and from that point forward all eligible applicants will be offered the program. RochesterWorks! and AAWDC will instead continue to enroll participants into both the program and the study throughout the life of the grant (i.e., October 2018). We project participant sample sizes of 1,220 at RochesterWorks!, and 1,800 at AAWDC, each split evenly between treatment and control participants. Overall, we project 5,020 study participants for the RTW Evaluation across the four grantee sites.

The follow-up survey will be administered to all 5,020 sample members. If all sites enroll individuals at the level planned, all sample members randomly assigned will be part of the respondent universe. If sites enroll more individuals than projected, those randomly assigned after the site-specific target is met will not be included in the survey (RochesterWorks! and AAWDC only). Exhibit B.1 shows expected sample sizes and predicted response rates for each of the grantees included in the evaluation.

[1] At all four sites, veterans are ineligible for the evaluation, and are excluded from random assignment. Other eligibility requirements vary by grantee.

Exhibit B.1 Projected Sample Sizes and Response Rates for 18-month Follow-up Survey

Grantee	Expected Sample Size	Predicted Number of Surveys Returned	Expected Response Rate
AAWDC	1,800	1,440	80%
JVS	1,000	800	80%
RochesterWorks!	1,220	976	80%
Worksystems, Inc	1,000	800	80%
Total	**5,020**	**4,016**	**80%**

B.2 Procedures for Collection of Information

B.2.1 Sample Design

No sampling will be required for the RTW 18-month follow-up survey.

B.2.2 Estimation Procedures

We start this section with a restatement of the RTW Evaluation research questions outlined in Section A.1.1 of Support Statement A. Separately for each program, what is the impact on:

1. Individuals' participation in education and training services?
2. The range of supports received, specifically receipt of advising, financial aid assistance, and employment assistance?
3. Educational attainment, including the receipt of credentials from training? To what extent do individuals progress to the next step of training on a career ladder?
4. Participants' ability to work, such as their attitudes towards work and self, and access to transportation or childcare?
5. Employment and earnings?
6. The characteristics of jobs, including wages, work hours and schedule, benefits, and sector of employment?
7. Participants' total household income and receipt of public assistance benefits?

The aim of the study is to estimate net impact: outcomes for individuals with access to the RTW program, relative to what outcomes would have been without the program, holding all else equal.

Our approach builds on the random assignment design of the evaluation. With random assignment, outcomes for the control group—those randomly chosen *not* to be offered program services—are a valid proxy for what the outcomes *would have been* for the treatment group, if they likewise had *not* been offered the program. It follows that comparing observed outcomes for the treatment and control groups provides an unbiased estimate of the causal effect of being offered the given RTW training program.

This evaluation will estimate program impacts for each of the areas specified in the research questions separately for each of the four grantees. Because this is a random assignment study, a simple comparison of mean outcomes for treatment and control participants for a given RTW grantee will yield valid

estimates of the causal impact of being offered the given RTW training regime. Nevertheless, estimating impacts via linear regression on baseline characteristics among the sample participants will yield more precise estimates of site-specific effects:

$$y_{s,i,t} = \alpha_s + D_{s,i}\delta_s + X_{s,i}\beta_s + \tau_t + \varepsilon_{s,i,t}. \qquad [B.1]$$

In Equation B.1, $y_{s,i,t}$ is the outcome variable (e.g., earnings) for respondent i from site s observed at time t, which is modelled as potentially varying with whether she is offered treatment $D_{s,i}$ (equal to 1 if i is a treatment participant, or 0 if she is a control), her background characteristics $X_{s,i}$ (measured prior to randomization), the point in calendar time τ_t, and an idiosyncratic time-specific error $\varepsilon_{s,i,t}$ (assumed to be on average equal to zero). The parameter of interest is δ_s, the level difference in outcome y for those participants who are offered the treatment ($D = 1$) at site s, relative to those who are not ($D = 0$), adjusting for (given random assignment, random) differences in pre-randomization characteristics.

Note that α_s allows for systematic differences in the level of earnings across sites s; e.g., earnings are much higher in San Francisco, CA than in Rochester, NY. Likewise, the subscript s on β_s allows for the possibility that the relationship between baseline characteristics X and outcome y may vary by site. For instance, the returns to education may be stronger in one local economy than in another. Most importantly, δ_s allows for variation in the impact of the RTW-funded services across sites s.

The evaluator plans to report estimates of impact relative to the control group. As is standard practice in the analysis of random assignment data, the team will use linear regression as the main estimation approach both for continuous outcomes (e.g., earnings or hours worked) and for binary outcomes (e.g., any employment).[2, 3] In particular, the evaluation team will use weighted least squares for outcomes measured in the follow-up survey, using weights to adjust for survey non-response. In addition, the evaluator will compute robust standard errors to correct for heteroscedasticity, both in general and as induced by binary outcomes (Winship and Radbill, 1994). The evaluator will report full sets of regression results—the nuisance parameters α_s and β_s to be estimated from Equation B.1—for selected key outcomes.

B.2.3 Degree of Accuracy Required

Exhibit B.2 provides a power analysis for possible site-specific sample sizes for the RTW Evaluation. Specifically, it reports minimum detectable impacts (MDIs) for binary outcomes, such as receipt of a degree, certificate, or credential, for various potential sample sizes and baseline rates (i.e., observed average outcomes among control group participants). MDIs are the smallest true impacts that the study has at least an 80-percent probability of detecting as statistically significantly different from zero.

As site recruitment is ongoing and survey response rates are unknown, final survey sample sizes are not currently known. Consider, however, the case of a site with 1,000 participants (as projected for JVS and WSI) – for which we expect to collect 800 survey responses (an 80% response rate), and a baseline rate of 25 percent. (In recent evaluations of similar training programs, the control group level of "received an educational or training certificate" was roughly 25 percent.) Under such conditions, Exhibit B.2 shows that we will have an 80 percent probability of detecting a true impact of 8.0 percentage points as

[2] As a robustness test, we also propose to use logistic models for binary outcomes to ensure that the signs and significance of the estimates from the linear and logistic models are the same.

[3] As is standard practice, we use ordinary least squares (the linear probability model) even for dependent variables that are bounded, such as hours and earnings (bounded below at 0).

statistically significant at p<0.05.[4] For a binary outcome such as completion of an educational or vocational training certificate, this MDI implies that our proposed design can detect impacts much smaller than those found in a recent evaluation of a similar DOL-funded job training program, conducted by Abt Associates.

Exhibit B.2 Minimum Detectable Impacts (MDIs) for Binary Outcomes

N = T + C		Potential Baseline Rate: (proportion observed among control group participants)			
Participant Sample Size	Expected Survey Responses	10%	25%	33%	50%
600	480	7.2 p.p.	10.4 p.p.	11.3 p.p.	12.0 p.p.
800	640	6.2 p.p.	9.0 p.p.	9.8 p.p.	10.4 p.p.
1,000	800	5.6 p.p.	8.0 p.p.	8.7 p.p.	9.3 p.p.
1,200	960	5.1 p.p.	7.3 p.p.	8.0 p.p.	8.5 p.p.
1,400	1,120	4.7 p.p.	6.8 p.p.	7.4 p.p.	7.8 p.p.
1,600	1,280	4.4 p.p.	6.4 p.p.	6.9 p.p.	7.3 p.p.
1,800	1,440	4.2 p.p.	6.0 p.p.	6.5 p.p.	6.9 p.p.

Notation: "p.p." stands for percentage points.

T/C Balance: We assume balanced assignment (i.e., 50% treatment, 50% control).

Standard Power Parameters: 80% power (*B*), α=0.05, two-tailed statistical tests.

Covariates/R^2: We estimate that the inclusion of sample members' baseline characteristics as covariates in the impact regressions will account for approximately 20 percent of the total variation in individual outcomes (R^2 = 0.2).[5]

Design Effect (*Deff*): Based on recent Abt experience with similar populations, we assume a design effect of 1.1.

MDI Calculation:

$$MDI = M\left(\frac{\left(1-R^2\right) Deff\, 2\left(p\left(1-p\right)\right)}{N/2}\right)^{1/2}$$

, where

- *M* is a function of the Z-score of the assumptions above (α=5%, thus $\alpha/2$=2.5%; *B* = 80%, thus 1-*B* = 20%; $M = Z_{2.5\%} + Z_{20\%}$ = 1.96 + 0.84 = 2.80),
- *p* is the baseline rate in the given column (e.g., 10%), and *p(1-p)* is the corresponding variance of the given binary outcome (e.g., receipt of degree), and

N is the sample size (expected number of survey responses).

Notes: This exhibit reports minimum detectable impacts (MDI), in terms of percentage points (p.p.), for a binary outcome, for a range of sample sizes and baseline rates. As noted elsewhere, we expect the follow-up survey to achieve an 80 percent response rate. The MDIs reported are for detection of improvements, but are capable of detecting reverses. The assumptions and calculations are similar to those in Abt Associates (2014) for the evaluation of Pathways for Advancing Careers and Education (PACE) (OMB No. 0970-0397).

B.2.4 Who Will Collect the Data and How Will It Be Done

Abt SRBI will administer the 18-month follow-up survey by telephone. The interviews are conducted by professional interviewers working in a centralized computer-assisted telephone interview (CATI) system that allows real-time error checking and observation by supervisors. If a person cannot be contacted by phone, Abt SRBI will attempt to conduct the survey in-person. The evaluation team estimates a response rate of 80 percent.

[4] Note that mathematically, the power calculations are symmetric for a control group mean *q* (e.g., 33%) or (1-*q*) (e.g., 67%). Thus the MDI for a given sample size (row) in Exhibit B.2 will be identical for a baseline rate of 33 percent and 67 percent, 25 percent and 75 percent, and 10 percent and 90 percent.

[5] The projected variance reductions due to use of baseline variables are from Nisar, Klerman, and Juras (2013).

B.2.5 Unusual Problems Requiring Specialized Sampling Procedures

Not applicable.

B.2.6 Periodic Data Collection Cycles

The 18-month follow-up surveys will be administered once. Building on experience conducting follow-up surveys with similar populations, the evaluator is implementing pro-active tracking of study participants between the time they are randomly assigned and the follow-up survey. These efforts are intended to update study participant contact information. Abt SRBI will send participants an email or text reminder for this purpose approximately every three months between sample enrollment and survey administration. Doing so will ensure that the researchers can effectively and efficiently contact study participants for the 18-month follow-up survey.

B.3 Methods to Maximize Response Rates and Address Non-response

The methods to maximize response rates are discussed with regard first to participant tracking and locating, and then regarding the use of monetary tokens of appreciation.

B.3.1 Participant Tracking and Locating

The RTW Evaluation team developed a participant tracking system, in order to maximize response to the 18-month follow-up survey. This includes using both active outreach to sample members and passive tracking resources. The active tracking planned for the RTW Evaluation begins with a welcome packet, sent to all sample members approximately one month after enrollment. This packet will consist of a welcome letter, a study brochure, and a $2 bill. The welcome letter and study brochure provide information about the tracking and survey data collection activities, and provides respondents with the option of updating their contact information, as appropriate. Additionally, Abt SRBI will send a text or email approximately once every three months following random assignment to remind sample members of their participation in the study.

In addition to the direct contact with participants, Abt SRBI will conduct several database searches to obtain additional contact information. Passive tracking resources are comparatively inexpensive to access and generally available, although some sources require special arrangements for access.

B.3.2 Tokens of Appreciation

Offering appropriate monetary gifts to study participants in appreciation for their time can help ensure a high response rate, which is necessary to ensure unbiased impact analysis. Those who complete the 18-month follow-up survey will receive a money order or gift card for $25 as a token of appreciation for their time spent participating in the survey.

B.3.3 Sample Control during the Data Collection Period

During the data collection period, the research team will minimize non-response levels and the risk of non-response bias in the following ways:

- Using trained interviewers (in the phone center) who are skilled at working with the sample population and skilled in maintaining rapport with respondents, to minimize the number of break-offs, and thus the incidence of non-response bias.

- Using an advance letter that clearly conveys the purpose of the survey to study participants, the incentive structure, and reassurances about privacy, so that they will perceive that cooperating is worthwhile.
- Using updated contact information captured through the contact information updates or the tracking communications conducted quarterly to keep the sample member engaged in the study and to enable the research team to locate him or her for the follow-up data collection.
- Taking additional tracking and locating steps, as needed, when the research team does not find sample members at the phone numbers or addresses previously collected.
- Employing a rigorous telephone process to ensure that all available contact information is utilized to make contact with participants.
- Administering the survey in person in instances where the participant cannot be surveyed by phone.
- Requiring the survey supervisors to manage the sample in a manner that helps to ensure that response rates achieved are relatively equal across treatment and control groups and sites.

The researchers will link data from various sources through a unique study identification number. This will ensure that survey responses are stored separately from personal identifying information, thus ensuring respondent privacy.

B.3.4 Nonresponse Bias Analysis and Nonresponse Weighting Adjustment

If, despite our best efforts, the response rate in a site is below 80 percent, the evaluator will conduct a nonresponse bias analysis. Regardless of the final response rate, the evaluator will construct nonresponse adjustment (NRA) weights. Using both baseline data collected just prior to random assignment and post-random assignment administrative data from the NDNH, the evaluator will estimate response propensity by a logistic regression model. Within the combination of site and experimental arm, study participants will be allocated to nonresponse adjustment cells defined by the intervals of response propensity. Each cell will contain approximately the same number of study participants. Within each nonresponse adjustment cell, the empirical response rate will be calculated. Respondents will then be given NRA weights equal to the inverse empirical response rate for their respective cell.[6] The use of nonresponse adjustment cells typically results in smaller design effects. The number of cells will be set as a function of model quality (five is a conventional value). The empirical response rates for a cell should be monotonically related to the average predicted response propensity. The evaluator will start with a large number of cells and reduce that number until it obtains the desired monotonic relationship.

Once provisional weights have been developed, the evaluator will look for residual nonresponse bias by comparing the estimates of the effects of the RTW-funded services on outcomes measured in the NDNH administrative data (which should be available for all study participants), estimated with the NRA weights in the sample of survey respondents vs. the estimates of the same effects estimated on the entire randomized sample (including survey nonrespondents) without weights. If they are similar (e.g., within each other's confidence intervals), then Abt will be reasonably confident that it has ameliorated nonresponse bias. If, on the other hand, there are important differences, then Abt will search for ways to improve our models and recalculate the weights as in Judkins, et al. (2007). [7]

[6] An alternative propensity adjustment method could use the directly modeled estimates of response propensity. However, these estimates can sometimes be close to zero, creating very large weights, which in turn lead to large survey design effects.

We last discuss here the evaluation's approach to missing covariate data. A dummy variable adjustment approach will be used to address item-non-response in the baseline covariate data. This strategy sets missing cases to a constant and adds "missing data flags" to the impact analysis model. This approach is easy to implement, and Puma et al. (2009) show that it works well for experimentally-designed evaluations. As detailed by Puma et al., the dummy variable adjustment approach involves the following three steps:

1. For each baseline covariate *X* with missing data, create a new variable *Z* that is set equal to *X* for all cases where *X* is non-missing, and set to a constant value for those cases where *X* is missing.
2. Create a new "missing data flag" variable *D*, which is set equal to one for cases where *X* is missing and set equal to zero for cases where *X* is not missing.
3. In the impact analysis model use *Z* and *D* (not *X*) as baseline covariates. This allows for the impact model to estimate the relationship between *Y* and *X* (via *Z*) when *X* is not missing, and to estimate the relationship between *Y* and *D* when *X* is missing.

B.4 Tests of Procedures or Methods to be Undertaken

The 18-month follow-up survey questions have been drawn from: (1) the 18-month follow-up survey for the Green Jobs and Health Care (GJ-HC) Impact Evaluation, conducted for DOL (OMB # 1205-0506); (2) the 36-month combined follow-up survey for the Pathways for Advancing Careers and Education (PACE) Evaluation and the Health Profession Opportunity Grants (HPOG) Evaluation, for the Administration for Children and Families (ACF) at the Department of Health and Human Services (OMB #0970-0394); (3) the 15-month follow-up survey for the PACE Evaluation (OMB # 0970-0397) for ACF; (4) the Baseline Information Form (BIF) for the RTW Evaluation (OMB # 1205-0507); and (5) the 15-month follow-up survey for the Workforce Investment Act (WIA) Adult and Dislocated Worker Program Gold Standard Evaluation, developed by Mathematica Policy Research for DOL (OMB #1205-0504). There are also several questions that were newly-developed for this survey. It should be noted that some of the items used in the PACE and combined PACE/HPOG surveys–questions on attitudes toward work and self–are drawn from questions and scales used in even earlier studies.

Together, the estimated length of the survey is approximately 40 minutes. This estimate is drawn from the time required to field the relevant portions of the other surveys listed in the previous paragraph. Additionally, we will conduct a formal pretest of the follow-up surveys, with a convenience sample of no more than nine respondents, with characteristics comparable to the study participants. These pretests will provide more definitive estimates about the length of the survey and their various components, as well as lead to improvements in questions, introduction scripts, wording, and document formatting. Following the pretests, respondents will be debriefed about the clarity of the questions and any potential problems with the instruments. The pretest findings will be used to modify the instrument as needed. However, given that many of the questions are from existing surveys, we do not expect many changes in the instrument after piloting.

[7] Judkins, et al. (2007) showed how to perfectly balance respondents and nonrespondents in a limited number of dimensions using a procedure that is call "weight raking." Using these weights, tabulations of respondents agree perfectly with tabulations based on the entire sample. This has been demonstrated to work on as many as about a dozen categorical variables at a time.

B.5 Individuals Consulted on Statistical Aspects of the Design

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

DOL

Gloria Salas-Kos	Contracting Officer's Representative, Employment and Training Administration
Molly Irwin	Chief Evaluation Office

Abt Associates

Karin Martinson	Project Director
Jacob Klerman	Director of Analysis
David Judkins	Statistician
Jane Leber Herr	Analyst
Kelly Daley	Survey Operations (Abt SRBI)

B.6 Individuals Responsible for Data Collection and Analysis

The following individuals are responsible for data collection and analysis for this study:

Abt Associates

Jacob Klerman	Director of Analysis
Jane Leber Herr	Analyst
Kelly Daley	Survey Operations (Abt SRBI)

References

Abt Associates. 2014. *Pathways for Advancing Careers and Education Evaluation Design Report*. OPRE Report #2014-76. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

CDC. 2012, July 27. Morbidity and Mortality Weekly Report. Supplement/Vol. 61, CDC's Vision for Public Health Surveillance in the 21st Century.

Judkins, D., D. Morganstein, P. Zador, A. Piesse, B. Barrett, and P. Mukhopadhyay. 2007. Variable Selection and Raking in Propensity Scoring. *Statistics in Medicine*, 26, 1022–1033.

Juras, Randall, Jacob Klerman, and Hiren Nisar. 2012. Comparing the Power of Designs for Evaluating Training Programs. Paper presented at ACF Welfare Research Conference, Washington, DC., June 2012.

Puma, Michael J., Robert B. Olsen, Stephen H. Bell, and Cristofer Price. 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 of Education Sciences, U.S. Department of Education. http://ies.ed.gov/ncee/pdf/20090049.pdf.

Winship, Christopher and Larry Radbill. 1994. Sampling Weights and Regression Analysis. *Sociological Methods and Research*, 23 (2), 230–257.