Supporting Statement for Paperwork Reduction Act Submission to OMB: Part B

Impact Evaluation of Upward Bound's Increased Focus on Higher-Risk Students

Cambridge, MA
Lexington, MA
Hadley, MA
Bethesda, MD
Chicago, IL

December 15, 2006

Prepared for
National Center for Education Evaluation and Regional Assistance
U.S. Department of Education
555 New Jersey Ave., NW
Washington, D.C 20208

Project Officer:
Jonathan Jacobson

Prepared by
Stephen Bell, Jeremy Luallen, Rob Olsen (Urban Institute), Ryoko Yamaguchi, Alan Werner
Abt Associates Inc.
55 Wheeler St.
Cambridge, MA 02138

Project Director:
Alan Werner

Abt Associates Inc.
55 Wheeler Street
Cambridge, MA 02138

Contents

B.   Collection of Information Employing Statistical Methods

B.1.   Respondent Universe

The evaluation is designed to measure the effects of the Upward Bound program, and it is designed to produce estimates that are nationally representative. To that end, the sampling plan for the evaluation has two stages: (1) the selection of a representative sample of grantees, and (2) the selection and random assignment of a representative sample of eligible applicants to Upward Bound. This section describes the universe from which the sample of grantees will be selected, the universe of students that the treatment and control groups selected for the evaluation are designed to represent, the random assignment process, and the samples we plan to select.

The universe of Upward Bound grantees from which the sample will be selected includes all 732 "regular" Upward Bound projects that are located within the contiguous United States. Excluded from the universe are Veteran's Upward Bound and Upward Bound Math Science projects, which provide a different set of services to different populations of students. Because there are limited numbers of Upward Bound grantees in Alaska, Hawaii, and the Pacific and Caribbean regions, and because the cost of including these grantees in the evaluation may be high, we have limited the universe to the contiguous States, including the District of Columbia.[1]

The target number of Upward Bound projects to participate in the evaluation is 90, since this number is needed to detect impacts for the 30 percent of students who must meet ED's definition of "higher-risk" (see Exhibit B.7 and related discussion). From the 732 regular Upward Bound projects in the specified universe, 100 projects will be randomly selected for the evaluation and an additional 20 randomly selected for a waiting list to be included only if needed. All 120 will be notified of their status early in 2007, contingent on OMB approval of the study. From the 100 projects selected as initial participants in the evaluation, we expect that about 10 will not be re-funded in the next fiscal year, leaving 90 projects eligible for the evaluation. If a larger proportion of the 100 projects is de-funded, or other grantees among the 100 cannot participate in the evaluation for some other reason—information that will become known in the first two months of recruitment—we will select projects from the waiting list in April, at random, to be added to the evaluation sample.

In its announcement for the grant cycle beginning in 2007, the Office of Postsecondary Education (OPE) within the U.S. Department of Education has specified that grantees selected for the evaluation must recruit twice as many applicants as there are openings in the 2007-2008 school year.   Under the evaluation plan, most eligible students who apply to enter Upward Bound during the 2007-2008 school year (including Summer 2007) will be subject to random assignment. (For more details, see Section B.2.1.) Based on the historical experience of Upward Bound projects, approximately 30 percent or more of a project's available slots open up each year. Using that proportion, as well as a strategy to sample grantees with a probability proportional to size (PPS), our estimates suggest that participating projects will have an average of 23 slots to fill via random assignment. If each project

---

[1]   The previous Upward Bound evaluation also restricted its sample of grantees to those in the contiguous States and District of Columbia.

recruits exactly twice as many eligible applicants as they have slots to fill, then we would expect 4,140 students (90 x 23 x 2) to go through random assignment in 90 projects. Half of these students would be assigned to the treatment group and offered the opportunity to enroll in the Upward Bound project to which they applied; the other half would be assigned to the control group and not permitted to enroll in a regular Upward Bound program.

From the approximately 4,140 students that will be randomly assigned in the 90 participating projects, 3,600 will be randomly selected for inclusion in the analysis of Upward Bound's impact based on data from a student follow-up survey.[2] As discussed in Section B.2.3 below, this sample is adequate to meet the needs of the evaluation. The plan for selecting the sample of 3,600 students from the more than 4,000 students included in random assignment is described in Section B.2.1

Exhibit B.1 shows the population of grantees and students served within the contiguous United States and the District of Columbia. With PPS sampling the distribution of grantees by state will tend to mirror the distribution of students in the population, ensuring that a wide range of states are part of the study sample.

Exhibit B.1

Population of Upward Bound Grantees and Students, by State

State	Grantees	Students Served	State	Grantees	Students Served
AL	36	2,547	NC	21	1,617
AR	16	1,182	ND	3	185
AZ	7	506	NE	6	398
CA	73	5,619	NH	2	156
CO	8	624	NJ	11	876
CT	5	350	NM	7	611
DC	7	516	NV	5	305
DE	5	323	NY	28	2,193
FL	21	1,391	OH	23	1,750
GA	18	1,528	OK	25	1,662
IA	17	1,165	OR	8	481
ID	4	301	PA	19	1,568
IL	28	1,997	RI	1	150
IN	10	731	SC	16	1,285
KS	12	749	SD	4	274
KY	18	1,337	TN	18	1,274
LA	18	1,485	TX	61	4,445
MA	15	1,092	UT	8	671
MD	11	823	VA	17	1,125
ME	6	452	VT	5	335
MI	20	1,644	WA	12	885
MN	18	1,265	WI	19	1,327
MO	16	967	WV	8	628
MS	8	709	WY	2	183
MT	6	406			

---

[2] We estimate that 90 percent, or 3,726, of the 4,140 students will provide consent to be in the study, of whom 97 percent, or 3,600, would be sampled for the impact analysis.

Population of Upward Bound Grantees and Students, by State

State	Grantees	Students Served	State	Grantees	Students Served
			Total	732	54,093

B.2. Procedures for the Collection of Information/Limitations of the Study

The lead contractor for the evaluation, Abt Associates Inc., will hire site liaisons to coordinate the data collection process at each Upward Bound site. Through the site liaisons, the study staff will coordinate with the staff of the Upward Bound programs that are participating in the study to obtain lists of eligible first-year program applicants. The site liaisons will work with the Upward Bound sites to distribute parent consent/student assent forms, the student selection forms, and the student baseline survey. Liaisons will also be responsible for collecting all of these items to ensure the data privacy. To obtain baseline information, a self-administered survey will be distributed to participants deemed eligible for the program during the program application process. In order to be the least burdensome to sites, the baseline survey will be incorporated into existing application materials and processes or administered by the site liaisons. Upon completion of the baseline surveys, students will seal them in precoded envelopes. Site liaisons at the Upward Bound sites will return completed surveys to Abt Associates.

B.2.1. Statistical Methodology for Stratification and Sample Selection

Below, we describe the sample selection plan, sample sizes, and the control group.

Sample Selection of Grantees
As noted previously, Upward Bound grantees funded for the 2007-2008 school year will be sampled on the basis of probability proportional to size (PPS). The number of funded slots will be used to define size. PPS sampling will be applied to strata defined by the type of host institution and co-location of other TRIO programs such as Talent Search, since these grantee characteristics may influence the effectiveness of Upward Bound:

- Type of host institution (4-year colleges and universities vs. projects hosted by two-year colleges and other institutions with less direct exposure to 4-year institutions)

- Co-location at the same host institution with other TRIO programs (either Talent Search —a less intensive TRIO program for Upward Bound-eligible students—or Upward Bound Math Science, which is similar in intensity to regular Upward Bound)

To increase power in detecting differences in impacts between groups and impacts on students served by certain subtypes of grantees, we plan to oversample projects that are co-located with Talent Search but not Upward Bound Math Science, which would otherwise encompass 24 percent of all Upward Bound students in the universe of study. Our power calculations (summarized in Exhibit B.6, below) suggest that it should be possible to detect impacts of Upward Bound for subgroups of 40 grantees, so we plan to sample 40 grantees co-located with Talent Search but not Upward Bound Math Science, in addition to 40 grantees co-located with *neither* Talent Search nor Upward Bound Math Science. Each

subgroup of grantees is of interest for a separate impact analysis, since the former offers the opportunity to estimate the impact of Upward Bound where Talent Search is readily available, whereas the latter offers the opportunity to estimate the impact of Upward Bound where Talent Search is less readily available. We plan to undersample projects co-located with Upward Bound Math Science, a stratum that encompasses 15 percent of all students in the universe, since control group students at these sites may have access to services similar in type and intensity to those available through regular Upward Bound, making these sites of lesser interest as a subgroup for estimating the effectiveness of particular types of Upward Bound programs.

As noted earlier, 100 projects will be sampled initially with the goal of achieving a final sample of 90 grantees that includes

- 40 projects co-located with Talent Search (but not Upward Bound Math Science)
- 10 projects co-located with Upward Bound Math Science
- 40 projects not in the first two groups.

When calculating impacts for the entire study population, or for subgroups of students drawn from all three of these strata, the data will be reweighted to match the distribution of students across grantee types that occurs in the population as a whole.

Sample Selection of Students
The plan for randomizing students in participating Upward Bound projects is designed to account for the differences across Upward Bound grantees in the timing of when students enter the program. Most Upward Bound applicants enter the program at the beginning of a session, such as the six-week summer session (a required program component in all Upward Bound projects), the fall semester, or the spring semester. However, while the most common entry point may be the start of the summer session at most projects, it could be the start of the fall or spring semesters in other projects. In addition, some students enter the program in the middle of a session, generally to fill a slot left vacant when a UB participant leaves the program.

To simplify the implementation of the evaluation and minimize burden on grantees selected for the evaluation, we plan to conduct "batch" random assignment for the applicants who apply to enter the program at a single entry point for each project. The single entry point will be selected separately for each participating grantee to maximize the study sample, but it will be constrained to be no later than the start of spring semester in 2008. We believe that under this plan, students who will enter the program during the 2007–2008 program year will generally be subject to random assignment.

Our evaluation plan involves setting the probabilities of being assigned to the treatment group in such a way to constrain any changes to the mix of students ordinarily served by the program and to obtain more complete cooperation from the grantees selected for the evaluation. Prior to random assignment, Upward Bound project directors will be asked to identify the eligible applicants that they would enroll in the program in the absence of the evaluation. The students that projects identify will be called "higher priority" students; other eligible applicants will be called "lower priority" students. In identifying these students, projects will be reminded of the new program requirements that 30 percent of new participants must meet a set of criteria associated with higher academic risk of failure.

For higher-risk students and students as a whole, projects will identify one-half of eligible students as higher priority students, and the other half as lower priority students.

To give Upward Bound project directors some control over the admissions process, and to provide them with an incentive to take the process of identifying higher priority students as seriously as they typically take the process of selecting students to participate in the program, the random assignment algorithm will assign two-thirds of the higher priority students and one-third of the lower priority students to the treatment group. As a result, for projects that receive exactly twice as many eligible applicants as they have slots to fill, the treatment group will include two higher priority students for every one lower priority student, and the control group will include two lower priority students for every one higher priority student. This approach gives each participating project the opportunity to prioritize students who meet different criteria without having to use complicated stratification schemes during random assignment that vary from project to project.

Exhibit B.2 and the accompanying table (Exhibit B.3) illustrate the random assignment process and results.
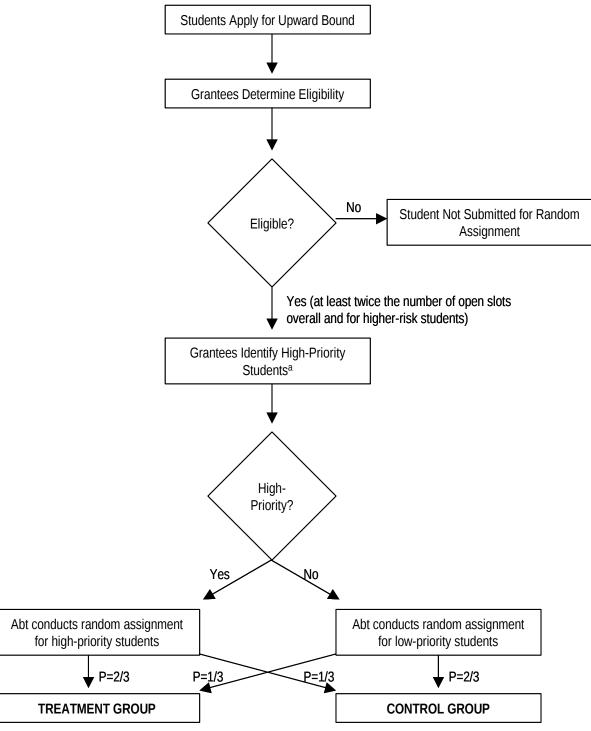
For smaller projects, we expect to include all randomized students in the target sample for the follow-up survey, which in total will encompass 1,800 treatment group students and 1,800 control group students. In these projects, the probability of selection into the treatment group will be twice as large for higher priority students as for lower priority students, and the probability of selection into the control group will be twice as large for lower priority students as for higher priority students. Analysis weights will be set equal to the inverse probability of selection for each subgroup to offset this imbalance in the realized sample.

The same randomization ratios will be used for larger projects, but the imbalance will be offset to some degree in selecting the follow-up survey sample (i.e., a greater share of the higher priority students randomized into the control group will be selected into the survey sample, offsetting their initially smaller number, while a greater share of the lower priority students randomized into the treatment group will be selected into the survey sample, offsetting their initially smaller number). For projects with at least 30 open slots to fill via random assignment this will produce self-weighted samples of survey respondents, since 10 of each type of student will be available for inclusion in the survey sample.[3] Any remaining imbalance in survey respondents across the four cells (high/lower priority x treatment/control)—including those expected for all projects with less than 30 openings to be filled through random assignment—will be removed through appropriate weighting of the analysis sample to equally represent the high and lower priority students in both the treatment and control groups. The examination of statistical power in section B.2.3 takes account of the uneven sample counts and the unequal analysis weights involved.

---

[3]   For example, consider a project with 60 eligible applicants for 30 open slots. The treatment group would consist of 20 higher priority students and 10 lower priority students, and the control group would consist of 10 higher priority students and 20 lower priority students. Inclusion of all members of each group of 10 individuals and random selection of 50 percent of each set of 20 individuals for the follow-up survey sample would produce a self-weighting sample of 10 higher priority students and 10 lower priority students in each experimental status.

Exhibit B.2

Student Intake and Random Assignment

```
                    ┌─────────────────────────────────┐
                    │  Students Apply for Upward Bound │
                    └─────────────────────────────────┘
                                   │
                                   ▼
                    ┌─────────────────────────────────┐
                    │  Grantees Determine Eligibility  │
                    └─────────────────────────────────┘
                                   │
                                   ▼
                              ◇─────────◇          No      ┌──────────────────────────────┐
                              │ Eligible? │ ──────────────▶│ Student Not Submitted for     │
                              ◇─────────◇                  │ Random Assignment             │
                                   │                       └──────────────────────────────┘
                 Yes (at least twice the number of open slots
                 overall and for higher-risk students)
                                   ▼
                    ┌─────────────────────────────────┐
                    │  Grantees Identify High-Priority │
                    │  Students[a]                     │
                    └─────────────────────────────────┘
                                   │
                                   ▼
                              ◇─────────◇
                              │  High-    │
                              │ Priority? │
                              ◇─────────◇
                           Yes           No
```

| Abt conducts random assignment for high-priority students | Abt conducts random assignment for low-priority students |

P=2/3   P=1/3   P=1/3   P=2/3

| **TREATMENT GROUP** | **CONTROL GROUP** |

a   To meet the 30-percent requirement, grantees would be advised to recruit twice as many higher-risk 9th graders as they intended to serve, and divide these students equally between the higher priority and lower priority groups.

Exhibit B.3

Random Assignment Overall and by Priority Group

	Type of Student	
	All	**Higher-Risk**
A. Random Assignment Overall		
Number of eligible students recruited by grantee	60	18
Number assigned to Treatment Group	30	9
Number assigned to Control Group	30	9
B. Random Assignment for High Priority Students		
Number of eligible students designated as "high priority"	30	9
Number assigned to Treatment Group (probability = 2/3)	20	6
Number assigned to Control Group (probability = 1/3)	10	3
C. Random Assignment for Low Priority Students		
Number of eligible students designated as "low priority"	30	9
Number assigned to Treatment Group (probability = 1/3)	10	3
Number assigned to Control Group (probability = 2/3)	20	6

Sample Size

As described above, we expect to include an average of 20 treatment group students and 20 control group students in the follow-up survey sample from each participating Upward Bound grantee (90 grantees in the evaluation). This results in a target sample for the survey of 3,600 students, equally divided between the treatment group (1,800 students) and control group (1,800 students). Of these, we expect to obtain completed follow-up interviews with 85 percent, resulting in a final analysis sample of 1,530 treatment group members and 1,530 control group members. However, small grantees may have fewer than 20 slots to fill and may not be able to contribute 40 students to sample. To adjust for this, we expect to select slightly more than 40 students from larger projects.

Control Group

Those students who are randomly assigned to the control group will not be accepted into the Upward Bound program at the time of randomization, and will be embargoed from participation in Upward Bound at any point.   During the course of the evaluation, the control group can participate in any other college preparation or other services, including other TRIO programs like Talent Search. We will take great care in collecting information from the control group on what relevant services they received during the two-year observation period.

B.2.2.   Estimation Procedure

The plans for the statistical analysis of the data are presented in Section A.16.

B.2.3. Degree of Accuracy Needed for the Purpose Described in the Justification

Minimum detectable effect sizes (MDESs)[4] have been calculated for the following populations of students:

- All students occupying funded slots run by the regular UB projects in the universe of interest

- The subset of those students who are at higher academic risk (assumed to be 30 percent of all students at each grantee)

- The subset of students given "highest priority" for admission (50 percent of all students in the research sample)

- The subset of students given "lower priority" for admission (50 percent of all students in the research sample)

- The subset of students at grantees with Talent Search grants but not Upward Bound Math Science grants (all students served by 40 grantees)

- The subset of students at grantees that are 4-year educational institutions (projected to be 51 grantees).

MDESs are calculated assuming:

- A two-tailed test for statistically significant impacts[5]

- alpha (probability of falsely rejecting the null hypothesis of 0 impact) = .05

- Power (probability of correctly rejecting the null hypothesis of 0 impact when an impact occurs) = .80

- Data for 85 percent of the students included in the follow-up survey (.85 x 3,600 = 3,060)

- Intra-class correlation, rho (i.e., share of total variance in student outcomes arising from variation in mean outcomes *between* projects, rather than variation in individual student outcomes *within* a project) = .022, .059, and .150 in different scenarios (see Exhibit B.4)

---

[4]   An effect size is defined as the impact of Upward Bound in the natural units of the outcome measure divided by the standard deviation of that measure.

[5]   It is sometimes argued that program evaluations need to look only for positive effects—and use one-tailed tests to do so—even if negative effects might occur, since the relevant policy question is "did the program improve outcomes." In such a framework, the null hypothesis of no effect is rejected only if strong evidence is found of *positive* impacts; evidence of negative impacts would be ignored and therefore need not be sought. One could also argue that negative effects are unlikely in the current study since a previous rigorous evaluation of Upward Bound discovered none that were statistically significant (using the two-tailed approach). However, the Upward Bound program has quite possibly changed considerably since then, and the implications of failing to find a positive impact might not be the same in the policy making process as finding a negative impact. Hence, a two-tailed testing approach seems the safer course. Were one-tailed testing adopted instead, all of the MDESs reported here would go down slightly.

- Variance of true impact across projects divided by variance in student outcomes among students, theta = .019, .041, and .075 in different scenarios (see Exhibit B.4)

- R^2 coefficient from regression explaining outcomes as a function of student-level covariates = .050, .130, and .150 in different scenarios (see Exhibit B.4).

Exhibit B.4

Parameter Values Assumed by Different MDES Scenarios

Scenario	Rho	Theta	R^2
Best case, empirically based[a]	.022	.019	.150
Cautious, empirically based	.059	.041	.130
Worst case	.150	.075	.050

[a] Sources for empirically-based estimates are "Statistical Power for Random Assignment Evaluations of Education Programs," Peter Z. Schochet, June 22, 2005, paper submitted to the Institute for Educational Sciences, U.D. Department of Education (used to derive theta) and a re-analysis of data from the previous Department of Education evaluation of Upward Bound (used to derive rho and $\underline{R}^{2)}$. The "best case" and "cautious" empirical scenarios represent the most and least favorable point estimates produced by those analyses. The third, "worst case" scenario makes even less favorable assumptions, relying on Schochet's recommended 'conservative' value to obtain theta and on customary values of rho and $\underline{R}^2$ from totally hypothetical minimum detectable effect size analyses done in studies that have no empirical data on which to rely.

All three sets of assumptions lead to minimum detectable effect sizes for impacts on all students well below 0.20 (see Exhibit B.5). Even in the worse case, an effect size of 0.16 will be detectable with 80-percent confidence for the study as a whole, and possible effect sizes as small as 0.11 in more hopeful scenarios. Given the high costs of the program per participant ($4,500), 0.20 standard deviations was determined to be a reasonable minimum detectable effects for the smaller subgroups.

Exhibit B.5

Impacts on All Students—Minimum Detectable Effect Sizes in Different Scenarios

Scenario	MDES
Sample size (follow-up survey N)	*3,060*
Best case, empirically based	.011
Cautious, empirically based	.012
Worst case	.016

Exhibit B.6 shows MDESs for analyses involving only a subset of the 90 grantees: impacts on students served by the 40 grantees in the sample with Talent Search grants (but no Upward Bound Math Science grants) and impacts on students served by the projected 51 grantees likely to be included in the sample that are 4-year educational institutions. These figures also indicate that effects

sizes of 0.20 or smaller will be detectable with 80-percent power in all but the worst-case scenario for projects co-located with Talent Search.[6]

Exhibit B.6

Impacts on Students Served by a Subset of Grantees – Minimum Detectable Effect Sizes

	MDES for Students Served by Grantees That…	
Scenario	**Have Talent Search Grants (40 grantees)**	**Are 4-Year Educational Institutions (51 grantees)**
Sample size (follow-up survey N)	*1,360*	*1,734*
Best case, empirically based	0.16	0.14
Cautious, empirically based	0.17	0.15
Worst case	0.23	0.20

MDESs for subgroups of students who are present at every grantee, such as those in the high-academic-risk target group, are shown in Exhibit B.7. Effects sizes of 0.20 or smaller for these subgroups will be detectable with 80-percent power in almost every scenario.

Exhibit B.7

Impacts on Subgroups of Students Served by All Grantees—Minimum Detectable Effect Sizes

	MDES for Students Who Are…		
Scenario	**Academically at High Risk (30% of total)**	**Higher priority for Admission (50% of total)**	**Lower Priority for Admission (50% of total)**
Sample size (follow-up survey N)	*918*	*1,530*	*1,530*
Best case, empirically based	0.20	0.16	0.16
Cautious, empirically based	0.21	0.16	0.16
Worst case	0.23	0.19	0.19

B.2.4. Unusual Problems Requiring Specialized Sampling Procedures

Not applicable.

---

[6]   Exhibit B.6 assumes that the variation in student outcomes and variation in true impacts across grantees is the same for the subgroups of grantees examined as for the sample as a whole. It is possible that both student outcomes and impacts will become more homogeneous as grantees of a particular type (e.g., those that are 4-year educational institutions) become the sole focus of the analysis; if so, MDESs will be smaller than shown.

B.2.5. Use of Periodic (Less Frequent Than Annual) Data Collection Cycles

The baseline survey, consent/assent forms, and student selection forms are one-time data collection efforts necessary to identify and describe students prior to random assignment.

B.3. Methods To Maximize Response Rates and Deal With Issues of Nonresponse

All eligible potential Upward Bound applicants will be required, before entering the admissions lottery, to return signed parental consent and student assent forms indicating whether they consent to be part of the study. Students will also be asked to complete the baseline questionnaire as part of their application process. We expect a 90 percent consent rate and a 90 percent response rate for the baseline survey. We expect a 100 percent response rate for the student selection form, since no student will be randomly assigned until the form is completed.

B.4. Tests of Procedures or Methods

In designing the baseline survey, we included items used successfully in previous studies or in national surveys. The survey questions have been thoroughly tested on large samples with prior OMB approval.

B.5. Names and Telephone Numbers of Individuals Consulted

The information for this study is being collected by Abt Associates Inc., Urban Institute, and Berkeley Policy Associates, research and consulting firms contracted to conduct the study on behalf of the Institute for Education Sciences (IES). With IES oversight, the contractors for the evaluation (Abt Associates Inc., Urban Institute, and Berkeley Policy Associates) are responsible of the study design, instrument development, data collection, analysis, and report preparation.

The instrument for this study and the plans for statistical analyses were developed by Abt Associates Inc. and the Urban Institute. The staff team is composed of Dr. Alan Werner, Project Director (Abt), Dr. Stephen Bell, co-Principal Investigator (Abt), and Dr. Robert Olsen, co-Principal Investigator (Urban Institute). In addition, the individuals listed below worked closely in developing the statistical procedures and will be responsible for data collection and data analysis. Contact information for these individuals is provided below.

Name	Title	Telephone
Alan Werner	Project Director, Abt	617-349-2832
Stephen Bell	Principal Investigator, Abt	301-634-1776
Robert Olsen	Co-Principal Investigator, Urban	202-261-5771
Ryoko Yamaguchi	Deputy Project Director, Abt	301-634-1778
Jeremy Luellen	Associate, Abt	617-349-2504
Jonathan Jacobson	IES Senior Research Scientist	202-208-3876
Marsha Silverberg	IES Economist	202-208-7178