Ticket to Work Program Evaluation Survey (National Beneficiary Survey)

OMB No. 0960-0666

Supporting Statement Part B

B. Collections of Information Employing Statistical Methods

1. Respondent Universe and Sampling Methods

Originally, we intended the NBS round 4 survey design to concentrate largely on following participants interviewed in earlier rounds and interviewing new Ticket Participants in Phase 3 states (see Table 1). The cross-sectional sample of Representative Beneficiaries in round 4 was planned to be substantially smaller than the cross-sections in earlier rounds. However, changes in the federal regulations which substantially alter the Ticket to Work Program mean that it is no longer meaningful to track long-term experiences of beneficiaries who participated in the program under the old regulations. As a result, we have made three necessary changes to the sample design for the round 4 survey: 1) eliminated the longitudinal TTW participant follow-up and substituted a representative cross-sectional participant sample, 2), changed the stratification for the TTW participant sample (from EN payment type to provider type), and 3) increased the size of the Representative Beneficiary Sample. Mathematica Policy Research (MPR), the data collection contractor for the three prior rounds, will modify the sample design, revise the questionnaire, conduct the round 4 data collection, and prepare data files and documentation. Table 2 shows how the revised sample compares to the original allocation.

TABLE 1
original NATIONAL BENEFICIARY AND TTW PARTICIPANT SAMPLE SIZES

Samplea		Round 1	Round 2	Round 3	Round 4	All Roundsc
Representative Beneficiary Samples		7,200	4,800	2,400	1,500	15,900
Longitudinal TTW Participant Samples	Phase 1 Cohorts (1)b	1,000	922	850	784	3,556
	(2)		1,000			1,000
	Phase 2 Cohorts (1)		1,000	922	850	2,772
	(2)			1,000		1,000

	Phase 3 Cohorts (1)			1,000	922	1,922
	(2)				1,000	1,000
	Total	1,000	2,922	3,772	3,556	11,250
Total Sample Size		**8,200**	**7,722**	**6,172**	**5,056**	**27,150**

Source: NBS Sample Design Report (Bethel and Stapleton 2002).

a Sample sizes refer to number of completed interviews
b(1)=TTW participant longitudinal sample and (2)=TTW participant cross-sectional supplement
c The All Years column is a tabulation of the number of interviews, not the number of sample members. Longitudinal cases may be included up to three times in these counts, depending upon the number of completed interviews for the sample member in question.

TABLE 2
ORIGINAL NBS AND REVISED ROUND 4 SAMPLE

Samplea	Original Round 4 Sample	Revised Round 4 Sample
Representative Beneficiary Sample	1,500	2,400
Ticket Participant Sample	3,556	3,000
Total	5,056	5,400

a Sample sizes refer to number of completed interviews

1.1. Universe and Sample

The universe for the round 4 NBS is the population of SSI and SSDI beneficiaries who meet the following criteria:

- Are between the ages of 18 and full retirement age;
- Are in current pay status based on SSA records (receiving either SSI and/or SSDI benefits) during the month in which the sample is drawn;
- Are not members of either of the two small groups of beneficiaries who will be ineligible for TTW;[1] and
- Are not non-disabled dependents of DI beneficiaries.

The survey sample has two major components: (1) the Representative Beneficiary Sample, and (2) the Ticket-to-Work Participant Sample. For round 4 of the NBS, we will use the most current counts of SSA Representative Beneficiaries and of Ticket Participants in the TTW program (see Table 3). The current population of SSI and SSDI beneficiaries ages 18 to current retirement age in current pay status is approximately 12,260,000 persons.

TABLE 3
Ticket Assignments as of June 2009

Tickets Assigned to ENs	Tickets Assigned to VRs	Tickets Assigned to In-Use VRs	Total
25,029	12,136	233.403	270,568

[1] Beneficiaries who are designated as Medical Improvement Expected (MIE) at award before their first Continuing Disability Review, and those who are eligible as children that have not completed their adult (age 18) re-determination.

1.1.1. Representative Beneficiary Sample

For the Representative Beneficiary sample, the target population will include beneficiaries in all 50 states and the District of Columbia. We expect two subpopulations of beneficiaries to be ineligible for Ticket assignment:

1. Beneficiaries who were designated as Medical Improvement Expected (MIE) at the time they received their allowances, and who have not yet completed a first Continuing Disability Review (CDR).

2. Young SSI recipients who are receiving benefits because of their eligibility as a child (younger than 18 years), and are in the process of completing a re-determination under the adult eligibility criteria.

Although these beneficiaries are not eligible for Ticket participation, they will be included in the survey samples to give complete coverage of the National Beneficiary population.

The Representative Beneficiary sampling frame will be selected from the SSA files, Beneficiaries will be selected into the frame if they satisfied the age requirement (18 years of age to full retirement age) in a pre-defined month and meet the active pay status criteria used in prior rounds.

In order to ensure a sufficient number of persons seeking work, as in prior rounds, the Representative Beneficiary Sample will be classified into sampling strata based on age, with persons in the younger age categories selected at higher rates than persons in the oldest age category. The Representative Beneficiary Sample will be divided into the following age groups, 18-24, 25-39, 40-49, and 50 and older, which will be used as the sampling strata. The target number of completed interviews for round 4 is 667 beneficiaries in each of the three younger age groups (18-24, 25-39, and 40-49). For the 50 and older age cohort, the target number of completed interview is 400 beneficiaries. The total target completes is 2,400.

1.1.2. Ticket Participant Sample

For the Ticket participants, the target population will consist of beneficiaries who newly assigned their ticket after the implementation of significant changes in July 2008. Ticket Participant sample members will be pulled from SSA files. Only participants who have used a ticket at least once in 2009 will eligible for the frame. Only deceased cases will be removed from the participant frame.

For Ticket Participants, the target sample will be based on the service provider types: traditional State Vocational Rehabilitation Agencies (SVRA) and the ENs created under TTW. Participant sample members will be selected through a multi-stage sampling frame of TTW participants who began TTW program participation after the re-launch of the Ticket program in 2008. The target number of completed interviews is 750 participants who have

used a traditional SVRA and 2,250 participants with tickets assigned to ENs (1,500 assigned to non-SVRA ENs, and 750 assigned to SVRAs acting as ENs).

1.2. Response Rates

SSA expects to attain a response rate of at least 80 percent for both samples. These targets are based on the contractor's extensive experience with prior rounds of the NBS. For round 3, the weighted response rate for the Representative Beneficiary Sample was 81.1 percent. The weighted response rate for the Ticket Participant Cross-Sectional Sample was 84.4 percent.

2. Procedures for the Collection of Information

2.1. Statistical Methodology for Stratification and Sample Selection

As originally proposed, SSA will use the same multi-stage clustered design developed for the first three rounds to facilitate in-person interviewing of beneficiaries who cannot be reached by telephone or who cannot be interviewed by telephone because of their disability or impairment. For the multi-stage design developed in prior rounds, data from SSA on the counts of eligible beneficiaries in each county was used to form the primary sampling units (PSUs) consisting of one or more counties. A stratified national sample of 80 PSUs was selected with two counties, Los Angeles County and Cook (Chicago) County, selected with certainty because of the number of SSA beneficiaries in these counties. Because of the size of these two counties (in both the beneficiary population and geographic size), secondary sampling units (SSUs) were formed using zip code of beneficiaries. Four and two SSUs were selected from Los Angeles county and Cook (Chicago) County, respectively.

For the fourth round, the sample of Representative Beneficiaries and Ticket Participants will selected from among beneficiaries residing in these same PSUs/SSUs. For the former, all beneficiaries in each area will be stratified into four age groups. The latter will be stratified based on service provider type.

2.2. Estimation Procedure

Most analysis will involve computation of descriptive statistics (means and percents) for the entire sample or specified sub-samples. We will use multivariate models (primarily multiple regression and probit or logit) in some instances.

The analysis of survey data from such complex sample designs requires the use of weights to compensate for variable probabilities of selection and special methods to compute standard errors. The base weight associated with a beneficiary who is sampled for the Representative Beneficiary Sample will be computed from the inverse of the selection probability. Based on the design developed for the TTW evaluation, the probability of selection is the product of the selection probability at each sampling stage: the PSU, secondary units (as needed), and the individual. Therefore, the initial sampling weight will

be the inverse of the full selection probability for each case. The calculation of the probability of selection is based on the following component probabilities:

1. The probability of selecting PSU *i* within PSU stratum *h*, *hi*, is *hi* = 1 for certainty PSUs; for non-certainty PSUs, the selection probability is given by

$$\pi_{hi} = n_h \frac{\text{MOS}_{hi}}{\text{MOS}_h},$$

where *nh* is the sample size for stratum *h*. Typically, *nh* = 1 or 2.

2. If secondary units are selected within the *hi*-th PSU, the probability of selecting secondary unit *j* is given by

$$\pi_{hij} = n_{hi} \frac{\text{MOS}_{hij}}{\text{MOS}_{hi}}.$$

where n_{hi} is the sample size for secondary units in PSU *hi*, MOS_{hij} is the measure of size of the secondary unit, and MOS_{hi} is the total measure of size for all secondary units in PSU *hi*.

3. When sub-areas are used, the probability of selecting a given beneficiary within stratum *s* of secondary unit *j* in the *hi*-th PSU is given by

$$\pi_{hijsk} = \frac{n_{hijsk}}{N_{hijsk}},$$

where *nhijsk* and *Nhijsk* are the sample and population size, respectively, for the *hijsk*-th stratum within secondary unit *j* of PSU *hi*, assuming sub-areas are used. When sub-areas are not used, *j* drops out of the subscripts.

Finally, the overall selection probability is given by the following:

$$\text{Overall selection probability} = \pi_{hi}\pi_{hij}\pi_{hijsk}.$$

The initial sampling weight is calculated as

$$\text{Base weight} = w_{hijsk} = \frac{1}{\pi_{hi}\pi_{hij}\pi_{hijsk}}.$$

The subscript *j* is dropped from the last two formulas for PSUs in which sub-areas are not sampled.

The use of base weights will yield unbiased estimates if there is adequate coverage and no non-response in the survey. Unit non-response (i.e., whole questionnaire non-response) occurs when an eligible sampled beneficiary fails to respond to the survey. To reduce the potential for bias due to unit non-response, the base weights will be adjusted with propensity scores, created using logistic regression models. Covariates in the logistic regression models

are variables that are available for both respondents and non-respondents, and are chosen because of their relation to likelihood of poor survey response and an assumed relationship to the data outcomes. At a minimum, candidates for covariates used in the logistic propensity models would include the strata used in sampling. It is important that each level of the model covariates has a sufficient number of sample members to ensure a stable adjustment. As with prior rounds, we will develop two logistic propensity models: (1) one for locating a person, and (2) one for response among located individuals. The models will be developed using data in the SSA database available on all sample members, which is extensive for most of the survey populations. The location and response logistic models will provide estimated propensity scores for each respondent that account for individuals with similar characteristics who could not be located or did not respond. The inverse of the propensity score will be used as the adjustment factor. The adjusted weight for each sample case will be the product of the initial sampling weight and the adjustment factor.

Propensity modeling can be viewed as the extension of the standard weighting class procedure. It will be used instead of the standard weighting class procedure because it allows us to use more factors (including both continuous and discrete factors) and more-complex interactions among factors to explain the differential propensity to be located or to respond. In addition, standard statistical tests are available to evaluate the selection of variables for the model. To identify the factors for inclusion in the models, we will use bivariate cross-tabulations and multivariate procedures, such as interaction detection procedures (for example, Chi-squared Automatic Interaction Detection, or CHAID, software). To evaluate the candidate factors and interactions, we will use a weighted step-wise procedure. The final model will then be checked by using survey data analysis software to obtain design-based precision estimates for assessing the final set of factors in the models. We expect that separate models may be required for some survey populations, because the factors explaining the ability to locate a person or response may be unique to these populations (for example, people who are TTW participants versus people in the National Representative Beneficiary sample).

After making adjustments for non-response, the contractor will further adjust the weights so that some weighted sample statistics match known population values. For example, if the weights (after adjustment to non-response) for recipients of SSI only, SSDI only, or both do not correspond to population values, the weights will be adjusted in a proportional fashion, so that the weighted sample and population values correspond. Potentially, the contractor can control to population statistics for any variable observed in SSA administrative data. The variables most likely to be used are Title, state, age, sex, months since award, and primary impairment.

In computing final weights, some individuals may end up with large weights. Variability in sampling weights can severely impact standard errors, particularly in the extreme case where one observation has a sampling weight that is orders of magnitude higher than other respondents. One procedure that will be used to alleviate this problem is “weight trimming.” In this procedure, the value of very large weights is simply reduced in magnitude, with the amount “trimmed” being distributed among other individuals in some way. Reducing the

weight can create biased estimates, but when one or two individuals have extremely large weights, the contribution to variance reduction outweighs the bias that might be created by trimming.

One way to protect against bias is to redistribute the "trimmed" amount over a group of individuals who share some common characteristic with those whose weights were trimmed. A characteristic believed to be important to key outcomes should be chosen if feasible.

As for prior rounds, we plan to create separate weights for the Ticket Participant Sample and the Representative Beneficiary sample, as well as a combined weight to represent the entire beneficiary population, including the Ticket Participant sample members.

2.3. Standard Errors

For the NBS, the sampling variance estimate is a function of the sampling design and the population parameter being estimated, and it is called the design-based sampling variance. The design-based variance assumes the use of "fully adjusted" sampling weights. The fully adjusted sampling weights are derived from the sampling design, with adjustments to compensate for locating a person, individual nonresponse, and ratio-adjusting the sampling totals to external totals (as described). We will follow the same method developed for prior rounds, developing a single fully adjusted sampling weight and information on analysis parameters (that is, analysis stratification and analysis clusters) necessary to estimate the sampling variance for a statistic, using the Taylor series linearization approach.

The Taylor series procedure is the most appropriate sampling variance estimation technique for complex sample designs such as the NBS. The Taylor series linearization procedure is based on a classic statistical method in which a nonlinear statistic can be approximated by a linear combination of the components within the statistic. The accuracy of the approximation is dependent on the sample size and the complexity of the statistic. For most commonly used nonlinear statistics (such as ratios, means, proportions, and regression coefficients), the linearized form has been developed and has good statistical properties. Once a linearized form of an estimate is developed, the explicit equations for linear estimates can be used to estimate the sampling variance. Because the explicit equations can be used, the sampling variance can be estimated using many features of the sampling design (for example, finite population corrections, stratification, multiple stages of selection, and unequal selection rates within strata). This is the basic variance estimation procedure used in SUDAAN, the survey procedures in SAS, STATA, and other software packages to accommodate simple and complex sampling designs. To calculate the variance, sample design information (such as stratum, analysis weight, and so on) is needed for each sample unit.

2.4. Degree of Accuracy Needed

2.4.1. Representative Beneficiary Sample

Given the new regulations which will change the structure of the TTW program, primarily by offering more generous payments to providers and new outreach to beneficiaries, boosting the Representative Beneficiary sample from that specified in the original design will provide information about how the new regulations affect interest in the TTW program, attitudes/interest in employment, the nature of current barriers, work activity; job characteristics, and the prevalence and nature of service use. Additionally, a larger cross-sectional sample of beneficiaries would allow cross round comparisons of changes in overall service use, employment, work attitudes, ticket awareness, and benefits.

A preliminary power analysis suggests that 2,400 completed cases will be sufficient to detect reasonable differences between rounds (see Table 4). In Table 3, the minimal detectable difference is a measure of the smallest difference between rounds that 2,400 completes in round 4 will be able to detect with 80 percent power and 90 percent confidence. For example, a minimal detectable difference equal to 3.8 percentage points indicates that if 10 percent of the beneficiaries were employed in round 1, and at least 13.8 percent of the beneficiaries were employed in round 4, the analysis would be able to detect a significant difference between round 4 and round 1. The minimum detectable differences would be only slightly larger for round 2 and round 4 comparisons and round 3 and round 4 comparisons, which would have smaller sample sizes.

TABLE 4

PROJECTED MINIMAL DETECTABLE DIFFERENCES BETWEEN ROUND FOUR AND
PREVIOUS ROUNDS

Round 4 Allocation	Round 4 and Round 1 (7,200 vs. 2,400)			Round 4 and Round 2 (4,800 vs. 2,400)			Round 4 and Round 3 (2,400 Each Round)		
	10%	25%	50%	10%	25%	50%	10%	25%	50%
Overall (100 Percent)	3.8%	5.5%	6.4%	4.0%	5.7%	6.6%	4.3%	6.2%	7.2%
Age 18 to 29	4.2%	6.0%	6.9%	4.4%	6.3%	7.3%	4.9%	7.1%	8.2%
Age 30 to 39	4.2%	6.0%	6.9%	4.4%	6.3%	7.3%	4.9%	7.1%	8.2%
Age 40 to 49	4.2%	6.0%	6.9%	4.4%	6.3%	7.3%	4.9%	7.1%	8.2%
Age 50 to 64	5.1%	7.4%	8.6%	5.4%	7.8%	9.0%	6.2%	8.9%	10.3%

The minimum detectable difference between two populations of an estimated percentage, $\hat{p}$, can be approximated by the following formula:

$$Var(\hat{p}) = P(1-P) \bullet (1/n_{eff1} + 1/n_{eff2}),$$

where $n_{eff\,1}$ and $n_{eff\,2}$ are the effective sample sizes of the two populations being compared and $n_{eff\,1} = n_1 / deff_1$ and $n_{eff\,2} = n_2 / deff_2$. The design effect is computed using the design effect due to unequal weighting and the design effect due to clustering, assuming 80 PSUs and an intracluster correlation of 0.02. The minimum detectable differences (using alpha = 0.05 and 80 percent power) is 2.80 square root (Var($\hat{p}$)).

2.4.2. TTW Participant Sample

Of the 3,000 completed TTW participant cases, 750 will be among participants who have used a traditional SVRA, 1,500 will be among participants with tickets assigned to non-SVRA ENs, and 750 will be among participants with tickets assigned to SVRAs acting as ENs. Therefore, subgroups of interest might be within a single provider type, with fewer than 750 participants. For this exercise, we assume a subgroup of 450 participants within a single provider type. With a sample of 450 participants with 80 PSUs, standard errors for estimated proportions of 10 percent, 25 percent, and 50 percent would be 1.54 percent, 2.24 percent, and 2.58 percent, respectively.[2] Applying the usual formulas for confidence intervals, that implies the following:

- Estimates of about 10 percent would have 95 percent confidence intervals of 7 percent to 13 percent.
- Estimates of about 25 percent would have 95 percent confidence intervals of 21 percent to 29 percent.
- Estimates of about 50 percent would have 95 percent confidence intervals of 45 percent to 55 percent.

2.4.3. Experiments

For the CATI/CAPI experiment, we will examine data comparability across modes including item non-response, number of options checked for check-all-that-apply items, non-differentiation among items in a series, proportion of agree/yes responses, length of responses to open-ended items, and distribution of responses or means for sensitive items. In addition, we will examine response effects by type of disability (cognitive or physical). To determine mode effects, cross tabulations will be developed to test the hypothesis of no association between mode for categorical variables. If the expected count in one or more cell in the contingency table is less than five, Fisher's Exact Test will be used rather than the chi-square statistic. Means will be calculated to test the hypothesis of no difference between

[2] This assumes a design effect of 1.2. This number was determined by reviewing the design effects obtained for key variables in subsamples of traditional, milestone and outcome, and outcome-only payment groups within Ticket participant samples from previous rounds. Under the assumption that dual sampling will not be necessary for any of the three service provider types, the design effect from the traditional subsample is a more realistic facsimile to the current situation. The design effect is computed using the design effect due to unequal weighting and the design effect due to clustering, assuming 80 PSUs and an intracluster correlation of 0.02. The minimum detectable differences (using alpha = 0.05 and 80 percent power) is 2.80 square root (Var($\hat{p}$)).

modes for continuous variables. A t-test will be used to determine whether a significant difference exists between modes for these items. The estimated number of completes that are needed for a minimum detectable treatment effect of .075 is 360 in the treatment group.

2.5. Unusual Problems Requiring Specialized Sampling Procedures

For the TTW participant sample, there is a chance that the number of participants within the PSU areas will be too small to meet sample size requirements. The number of participants with tickets assigned to ENs who are located in the PSUs may not be adequate. If this is the case, as in prior rounds, we will plan to use a hybrid design, which combines an unclustered stratified random sample with the clustered sample design. While both the unclustered and clustered samples are nationally representative, the data collection in the unclustered component will be limited to only CATI (no in-field follow-up or interviewing) due to the high cost that would be associated with field follow-up for the unclustered cases. For national estimates, we will compute sampling weights to account for this "dual-frame" strategy, as we have in the prior rounds.

The result will be lower response rates for the non-PSU participants, and potential bias in the estimates. To address the bias issue, the evaluator will compare the responses of the within-PSU phone interview sample to the responses of the within-PSU in-person interview sample. We expect the telephone response rate to be higher for participants than for all beneficiaries, because we know that these are individuals who are not being prevented from at least attempting to work by their physical or mental conditions, or by other personal circumstances.

2.6. Periodic Cycles to Reduce Burden

In a change from the original design, longitudinal TTW participants from prior rounds will not be included in this survey. We are substituting the follow-up of Phase 1, Phase 2, and Phase 3 longitudinal participants with a new cross-sectional sample of participants (3,000 completes) stratified on provider type. While the currently planned longitudinal follow-up samples in round 4 would have provided important information on the longer-term affects of TTW, we believe that the longitudinal data collected through round 3 will adequately address these issues and will reduce respondent burden.

3. Methods to Maximize Response Rates and to Deal with Nonresponse

3.1. Maximizing Response Rates

Locating sampled beneficiaries and participants is the first challenge in obtaining high response rates. While SSA has contact information for all beneficiaries, we know from the contractor's past experience this information does not always lead directly to the beneficiary. Telephone numbers can be particularly problematic, because there is no administrative reason to keep them updated in SSA records. Addresses are more reliable, because they are sometimes used for mailing checks. These might, however, be post office boxes, addresses of guardians, financial institutions, or other types of addresses that may make it difficult to locate the beneficiary. Further, since many beneficiaries now receive their checks via direct deposit, address information is less accurate than it once was. Contact

information for TTW participants is likely to be much better, because their EN needs the information to provide them with services, and will provide it to the evaluator via SSA.

To improve contact information, the contractor will mail an advance letter to each sampled person prior to the survey (see Appendix A), using the address of record (either from SSA administrative data or provider records). This letter will describe the survey and indicate that the beneficiary will be contacted regarding it. Locating will begin with letters that are returned to the contractor as undelivered. When an address is available without a phone number, the contractor will conduct a directory search to obtain a number. When director searches are unproductive, the contractor will submit searches to Accurint, a comprehensive database compiled from multiple sources; use locating letters, and telephone tracing (calling former neighbors or payees). The contractor located more than 90 percent of current beneficiaries for previous rounds.

If a phone number is available or obtained, the contractor will attempt to contact the beneficiary by telephone, to conduct the interview. The contractor will use protocol that calls for repeat attempts, including attempts on different days and different times. If successful contact is made and the beneficiary consents to be interviewed, the caller will conduct the interview, using CATI technology.

If telephone contact attempts are exhausted, or no phone number is available or obtained, the contractor will conduct a field search, starting with any available information, such as last address. CAPI interviewers, who have also been trained as field locators, will visit the former residences of beneficiaries and question neighbors, neighborhood businesses, likely relatives, and local post offices about the beneficiary's whereabouts. They will make inquiries of other local sources, such as the "corner store," libraries, churches, community centers, bars, and welfare offices. Delivering a note to the last known address that contains a toll-free number and the promise of an incentive payment can also be effective. Contacts with individuals other than the beneficiary will be handled carefully. Informants will be told the beneficiary is part of an important study on American life—no mention will be made of the beneficiary's SSI/DI status or health condition. Once contact information for the beneficiary is obtained, an interviewer will either phone the beneficiary and conduct a CATI interview, or will visit and attempt to conduct the interview using computer-assisted personal interview (CAPI) technology.

The impairments and health of some beneficiaries will make response problematic, especially by phone. To facilitate responses to the CATI interview, the contractor will offer the use of several assistive devices (amplifier phones, TTY phones, Telecommunications Relay Service, instant messaging, and sign interpreters for in-person interviews) and will instruct interviewers to remain patient, repeat answers for clarification, and complete the interview in more than one session if necessary. Despite these efforts, we know that some respondents will be unable to complete the interview by telephone; others will be unable to complete the interview at all.

If the respondent is unable or unwilling to undergo a phone interview, the interviewer will assess whether to interview a proxy for the respondent by telephone using an innovative "mini-cognitive test" designed expressly for the survey.[3]). If a telephone interview of a proxy respondent is either inappropriate or not feasible, or the beneficiary requests it, the evaluator will attempt to conduct an in-person interview. This interview might also be of a proxy respondent. To further improve the representativeness of the sample, we will allow for "assisted" interviews. These are different from proxy interviews because beneficiaries are present when the questions are asked, and they try to respond for themselves. The assistant, typically a family member, provides encouragement, interpretation, and verification of answers if needed.

A Spanish-language version of the instrument will also be developed and administered by Spanish-speaking interviewers to Spanish-speaking subjects. Translation services will be used for other non-English speakers.

As indicated in Part A, respondents will receive $10 to compensate them for their time, and will be assured of the confidentiality of their responses. A pre-paid incentive of $10 will be provided in the final weeks of the data collection period to encourage call-ins from non-respondents. These steps should also encourage sampled individuals to cooperate with the interviewer once contact is made.

3.2. Dealing with Issues of Non-Response

The contractor will adjust the base weights for non-response, using the procedures described above. They will also adjust the weights to control distributions for some variables to known totals from the administrative data, as described above. The evaluator can assess the extent of remaining bias by comparing weighted outcomes for the survey sample that can be observed in administrative data (e.g., annual earnings, SSI and DI payments, and Ticket payments) to outcomes for the population that the weighted sample is intended to represent. Such comparisons are expected to be especially important to assess attrition bias in analyses of the follow-up surveys for the longitudinal samples. The evaluator will also be able use the administrative data to assess the extent to which non-response in the follow-up surveys is due to mortality.

4. Tests of Procedures and Methods to be Undertaken

The original survey items were developed and initially pre-tested as part of a separate contract held by Westat. This testing involved two sets of cognitive interviews with a total of 12 beneficiaries, and two sets of pre-test interviews involving a maximum of nine interviews for each of the different groups of interest. After revisions were made by the current contractor to prepare the instrument for CATI/CAPI programming, another pretest

[3] Westat designed the test as part of the design of the Ticket to Work evaluation; MPR modified it after pretesting. MPR will conduct an assessment of interviewer reliability and accuracy in interpreting and coding responses both during interviewer training and throughout data collection.

was conducted to ensure that the instrument was clear and understandable to respondents and to test interviewer usability. The pretest sample was selected from beneficiaries and TTW participants who were not living in the sampled PSUs--74 pretest interviews were completed. Because the revisions made to the questionnaire for round 4 are relatively minor, and most questions have been administered in all three prior rounds, we do not plan to conduct additional testing.

5. Individuals Consulted on Statistical Aspects of the Design

We list the individuals consulted on technical and statistical issues related to the round 4 data collection in Table 5.

Table 5
individuals consulted on technical and statistical issues

Name	Affiliation/Address	Telephone Number
Paul O'Leary	Social Security Administration, Office of Research, Evaluation, and Statistics Washington, DC	(202) 358-6227
Eric Grau	Mathematica Policy Research	(609) 945-3330
Gina Livermore	Washington, DC 20024	(202) 264-3462
Frank Potter		(609) 936-2799
David Stapleton		(202) 484-4224
Craig Thornton		(202) 484-5277
Debra Wright		(202) 554-7576

References

Bethel, J. and D. Stapleton. “Evaluation Design For The Ticket To Work Program: Final Survey Sample Design.” Prepared for the Social Security Administration. Washington, DC 2002.