

SUPPORTING STATEMENT FOR

**THE NATIONAL BENEFICIARY SURVEY—
GENERAL WAVES AND SEMI-STRUCTURED INTERVIEWS**

OMB No. 0960-NEW

Supporting Statement: Part B

CONTENTS

| | |
|--|-----------|
| Part B: Collection of Information Employing Statistical Methods (see separate document) | 1 |
| B.1. Statistical Methodology..... | 1 |
| B.2. Procedures for the Collecting the Information..... | 4 |
| B.3. Methods to Maximize Response Rates..... | 11 |
| B.4. Tests of Procedures..... | 14 |
| B.5. Statistical Agency Contact for Statistical Information | 15 |
| Part B References | 16 |
| Part B Tables | |
| B.1 NBS–General Waves Sample Sizes by Strata..... | 3 |
| B.2 Projected Minimal Detectable Differences Between Groups In Representative Beneficiary Sample..... | 8 |
| B.3..Projected Minimal Detectable Differences Between Successful Worker Groups..... | 9 |
| B.4..Incentive Conditions..... | 10 |
| B.5..Individuals Consulted on Technical and Statistical Issues..... | 15 |

B. Collections of Information Employing Statistical Methods

B.1. Statistical Methodology

B.1.1. Respondent Universe and Sampling Methods

Originally, SSA intended the new NBS-General Waves survey design to include a national sample of SSA disability beneficiaries and a sample of beneficiaries whose benefits had been suspended due to successful work in all three rounds of data collection. We planned to complete approximately 4,000 interviews with active beneficiaries in each of the three rounds while the sample sizes for those whose benefits had been suspended would vary across rounds¹. In addition, some beneficiaries, identified as individuals who were in suspense status at the time of the round 1 interview, would be followed longitudinally in rounds 2 and 3.

However, due to difficulties associated with developing a sample design that would provide sufficient numbers of beneficiaries who earned enough to have had their benefits suspended in the recent past, SSA postponed the start of the survey to 2015 and will no longer include a sample of successful workers as part of the round 1 NBS-General Waves. In lieu of including a sample of successful workers in the survey data collection, we plan to conduct semi-structured interviews with this group. This will give us more time to settle on an adequate design for this sample so that they can be included at rounds 2 and 3 while still allowing us to collect important information about factors that aid or inhibit beneficiaries in their efforts to obtain and retain employment on the general sample. Mathematica Policy Research (MPR), the data collection contractor for the four prior rounds, will modify the sample design, revise the questionnaire, conduct all three rounds of the new data collection, and prepare data files and documentation.

NBS-General Waves - The primary purpose of the NBS-General Waves is to assess beneficiary well-being and interest in work, learn about their work experiences (successful and unsuccessful), and identify how factors such as health, living arrangements, family structure, pre-disability occupation, use of non-SSA programs (e.g., Food Stamps), knowledge of disability insurance (DI) and Supplementary Security Income (SSI) work incentive programs, obstacles to work, and beneficiary interest and motivation to return to work promote or restrict long-term work success. The NBS-General Waves will use a sample design similar to that used for prior NBS (conducted by SSA in 2004, 2005, 2006, and 2010).

We will field the first wave of the NBS-General Waves in 2015. We will conduct the subsequent two rounds in 2017(round 2) and 2019 (round 3). We will collect survey data not available from SSA administrative data or other source. The NBS is designed as a dual mode survey: data are collected primarily using CATI with CAPI available for those who request or require an in-person interview to facilitate their participation in the survey. The survey instrument will be identical in each mode. In all cases, interviews are attempted with the sample person. We will

¹ At round 1, we planned to interview approximately 4,500 beneficiaries whose benefits have been suspended due to work. In rounds 2 and 3, we planned to complete approximately 3,000 interviews with suspended beneficiaries that have been selected for the cross-sectional samples. In addition, 2,500 beneficiaries in suspense at round 1 would be followed longitudinally in rounds 2 and 3.

seek a proxy respondent only if the sample persons are unable to complete either a telephone or in-person interview as a result of their disability.

Semi-Structured Interviews - The primary purpose of the semi-structured qualitative interviews is to allow SSA to gain an in-depth understanding of factors that aid or inhibit beneficiaries in their efforts to obtain and retain employment and advance in the workplace. The qualitative data will add context and understanding when interpreting the survey results. The semi-structured interviews will also inform the sample and survey design of rounds 2 and 3 for the successful earners.

B.1.2. Universe and Sample

NBS-General Waves - The target population or “universe” for the NBS-General Waves will include all SSI or DI beneficiaries who meet the following criteria:

- Between the ages of 18 and full retirement age (FRA)-18 to 65 if receiving SSI, and 18 to 66 if receiving DI.
- In active benefit status² as of June 30, 2014 in either the SSI or DI program.
- Are not nondisabled dependents of DI beneficiaries.

To maintain consistency and support trend analyses, we will apply essentially the same sample selection criteria for the NBS-General Waves that we used to prepare the national samples in the prior NBS. In round 1, we will select only a nationally representative sample of active SSI recipients, DI and concurrent beneficiaries. In rounds 2 and 3, we will apply the same criteria but will add a stratum of SSI recipients and DI beneficiaries whom we have identified (using SSA administrative data) as having had high earnings from work in the year prior to the interview. The target population includes SSI recipients and DI beneficiaries in all 50 states and the District of Columbia. The estimated size of the target population is 13.3 million.³

Active Beneficiaries - In order to ensure a sufficient number of persons seeking work, the active recipient and beneficiary strata will be stratified into four age categories: 18 to 29, 30 to 39, 40 to 49, and 50 years old or over, as seen in Table B.1. We will select persons in the younger age categories at a higher rate than those in the oldest age category.

² Active status includes beneficiaries who are currently receiving cash benefits as well as those whose benefits have been temporarily suspended for work or other reasons. It does not include beneficiaries whose benefits have been terminated.

³ This figure was arrived at by comparing statistics from 2009 in the NBS round 4 Editing, Coding, Imputation, and Weighting Procedures report (Grau et al. 2011) to those in SSA Publication No 13-11785 from 2009 and 2011 (*Fast Facts & Figures About Social Security*, 2009, page 31 and *Fast Facts & Figures About Social Security*, 2011, page 31). The figures from the “Fast Facts” booklets are limited to beneficiaries between 18 and 64 years of age but include beneficiaries from the U.S. Trust Territories. The 2009 NBS includes unretired older workers not yet eligible for Social Security (older than age 64) but excludes beneficiaries from the U.S. Trust Territories.

Successful Workers - In rounds 2 and 3, we plan to add a successful worker stratum to the NBS. The successful worker stratum will be divided into three substrata identifying benefit type: SSI only, DI only, and those receiving SSI and DI (concurrent benefits). We will oversample cases from the successful worker stratum to ensure a large enough sample for analyses.

Table B.1. NBS–General Waves Sample Sizes by Strata

| Sampling Strata | Sample Size | Target Completed Interviews |
|---------------------------------------|-------------|-----------------------------|
| Round 1 | | |
| Active beneficiaries | 5,000 | 4,000 |
| Age range in years | | |
| 18 to 29 | 1,500 | 1,200 |
| 30 to 39 | 1,500 | 1,200 |
| 40 to 49 | 1,500 | 1,200 |
| 50 to FRA | 500 | 400 |
| Round 2 | | |
| Active beneficiaries | | |
| Age range in years | 5,000 | 4,000 |
| 18 to 29 | 1,500 | 1,200 |
| 30 to 39 | 1,500 | 1,200 |
| 40 to 49 | 1,500 | 1,200 |
| 50 to FRA | 500 | 400 |
| Successful Workers | 5,625 | 4,500 |
| SSI | 1,875 | 1,500 |
| SSDI | 1,875 | 1,500 |
| Concurrent | 1,875 | 1,500 |
| Round 3 | | |
| Active beneficiaries | 5,000 | 4,000 |
| Age range in years | | |
| 18 to 29 | 1,500 | 1,200 |
| 30 to 39 | 1,500 | 1,200 |
| 40 to 49 | 1,500 | 1,200 |
| 50 to FRA | 500 | 400 |
| Successful Workers | 3,750 | 3,000 |
| SSI | 1,250 | 1,000 |
| SSDI | 1,250 | 1,000 |
| Concurrent | 1,250 | 1,000 |
| Longitudinal successful worker sample | 2,813 | 2,250 |

The sampling design will include the selection of 80 primary sampling units (PSUs) along with selection of zip-code-based secondary sampling units (SSUs) within certainty PSUs. Specifically, we will select PSUs using a four-level composite size measure, incorporating the four age-based strata of the active national beneficiary sample.

Semi-Structured Interviews - At round 1, we will conduct semi-structured qualitative interviews with three groups of beneficiaries who have had periods of successful work: 1) those identified as having sustained high earnings from employment, 2) those identified as having high earnings from employment but who did not sustain high earnings, and 3) employed recipients and beneficiaries under the age of 30.

We will select interview participants through telephone recruitment efforts targeting beneficiaries and recipients SSA identifies (using administrative data) as having had high earnings or suspense status within the past 12 months. Some high earners may have had their cash benefits stopped, or suspended by SSA at the time of sample selection if their earnings exceeded the maximum threshold for receipt of cash benefits. In addition, we will attempt to recruit young employed recipients and beneficiaries under the age of 30; recipients and beneficiaries in this age group have the highest employment success rates. Once recruited, we will schedule an interview with the individual. As with the survey, we will provide accommodations for those with hearing or speech impairments. In addition, we will conduct qualitative interviews in both English and Spanish. Our goal is to conduct 20-30 interviews with each group of beneficiaries for a total of -no more than 90 current and past SSA beneficiaries. These sample sizes are based on widely accepted standards for qualitative interviewing which suggest that after 20-30 interviews, saturation is reached and little new information is obtained (Mason 2010). Statistical power is not applicable because this is a qualitative data collection.

B.1.3. Response Rates

NBS-General Waves - We expect to attain a survey response rate of 80 percent for both strata. These targets are based on the contractor's experience on studies with similar populations, including the Youth Transition Demonstration, Accelerated Benefits, and the prior NBS. We recognize that it is becoming increasingly challenging to locate sample members (especially with electronic payments) and to gain their cooperation with the survey process. If the response rate for both strata is less than 80 percent, we will conduct a non-response bias analysis and take the results into account during weighting procedures.

Semi-Structured Interviews - For the qualitative interviews, we expect to recruit 5 to 6 percent of DI beneficiaries and SSI recipients we contact regarding the interview, thus we will recruit from a list of approximately 1,800 current and past beneficiaries and respondents (identified as high earners in the recent past by SSA). We anticipate about 80 percent of respondents scheduled for interviews will complete interviews. We anticipate a high level of cooperation from because they will have 1) already been screened for eligibility and agreed to participate, 2) had an appointment scheduled, and 3) will have received a reminder phone call about the interview.

B.2. Procedures for the Collecting the Information

B.2.1. Statistical Methodology for Stratification and Sample Selection

NBS-General Waves - We will use the same multi-stage clustered design developed for the prior NBS to facilitate in-person interviewing of respondents selected for the NBS 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 the prior NBS, data from SSA on the counts of eligible DI beneficiaries and SSI recipients in each county were used to form 1,330 PSUs consisting of one or more counties. From this list, we will select a stratified national sample of 80 PSUs. As in the prior NBS, we anticipate that Los Angeles and Cook (Chicago) counties will be selected because of the number of SSA beneficiaries in these locations. Because of the size of these two counties (in both the beneficiary population and geographic size), SSUs were formed using beneficiaries' zip codes. Using the same set of SSUs created for the prior NBS, we will select four and two SSUs from the Los Angeles and Cook (Chicago) counties, respectively. We will select the PSUs and SSUs with probability proportional to size, where we define size as a composite size measure that accounts for the number of active beneficiaries and recipients in each age group.

Semi-Structured Interviews – We will use SSA administrative data to select a sample of SSI, SSDI, and concurrent beneficiaries and recipients who have achieved high earnings or suspense status in the prior year. We will use a telephone screener (see Attachment A) to identify them in three groups: 1) past high earners who have sustained high earnings from employment; 2) past high earners who have not sustained their high earnings; and 3) currently employed beneficiaries and recipients under the age of 30. We will identify and recruit until we obtain a sample of approximately 38 beneficiaries and respondents in each group (with the goal of completing up to 30 per group).

B.2.2. Estimation Procedure

NBS-General Waves - The analysis involves computation of descriptive statistics (means and percentages) for the entire sample or specified subsamples. Multivariate models (primarily multiple regression and probit or logit) will be used in some instances.

The analysis of survey data from such complex sample designs requires the use of weights to compensate for various probabilities of selection and special methods to compute standard errors. We will compute from the inverse of the selection probability the base weight associated with a DI beneficiary or SSI recipient who is sampled for the NBS-General Waves. The probability of selection is the product of the selection probability at each sampling stage—the PSU (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 , p_{hi} , is $p_{hi} = 1$ for certainty PSUs; for noncertainty PSUs, the selection probability is given by

$$\pi_{hi} = n_h \frac{MOS_{hi}}{MOS_h} ,$$

where n_h is the sample size for stratum h . Typically, $n_h = 1$ or 2 .

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

$$\pi_{hij} = n_{hi} \frac{MOS_{hij}}{MOS_{hi}} .$$

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 subareas are used, the probability of selecting a given beneficiary within stratum s of secondary unit j in the hi -th PSU is given by

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

where n_{hijks} and N_{hijks} are the sample and population size, respectively, for the $hijks$ -th stratum within secondary unit j of PSU hi , assuming subareas are used. When subareas are not used, j drops out of the subscripts.

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

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

The initial sampling weight is calculated as

$$\text{Base weight} = W_{hijks} = \frac{1}{\pi_{hi} \pi_{hij} \pi_{hijks}} .$$

The subscript j is dropped from the last two formulas for PSUs in which subareas are not sampled.

The use of base weights will yield unbiased estimates if there is adequate coverage and no nonresponse in the survey. Unit nonresponse (that is, whole questionnaire nonresponse) occurs when an eligible sampled beneficiary fails to respond to the survey. To reduce the potential for bias due to unit nonresponse, 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 nonrespondents, and are chosen because of their relation to the 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 will 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, the contractor will develop two logistic propensity models: one for locating a person and another for response among located individuals. We will develop the models 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 provide estimated propensity scores for each respondent that account for individuals with similar characteristics who are not located or did not respond. We will use the inverse of the propensity score as the adjustment factor. The adjusted weight for each sample case will be the product of the initial sampling weight and the adjustment factor.

We view propensity modeling 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 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. We will then check the final model using survey data analysis software to obtain design-based precision estimates for assessing the final set of factors. We expect 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 suspended from receiving benefits due to work versus people in current pay status).

After making adjustments for nonresponse, we will further adjust the weights so that some weighted sample statistics match known population values. For example, if the weights for recipients of SSI only, DI 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, we can control to population statistics for any variable observed in SSA administrative data. The variables most likely to be used are beneficiary type, 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. We will use “weight trimming” to alleviate this problem. 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. These “trimming classes” will be defined using variables selected in the same manner used to select variables for the nonresponse adjustments. Since we will use propensity modeling instead

of weighting classes to do the nonresponse adjustments, we will define trimming classes using the most important variables in the propensity models.

Semi-Structured Interviews - Due to the qualitative nature of the semi-structured interviews, we will not be calculating probability of selection for estimation purposes.

B.2.3. Standard Errors

NBS-General Waves - For the NBS-General Waves, the sampling variance estimate is a function of the sampling design and the population parameter being estimated; it is called the design-based sampling variance. The design-based variance assumes the use of “fully adjusted” sampling weights, which 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. We will follow the same method developed in the prior NBS, 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. It 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 and analysis weight) is needed for each sample unit.

Semi-Structured Interviews - Due to the qualitative nature of the semi-structured interviews, we will not be calculating standard errors.

B.2.4. Degree of Accuracy Needed

a. NBS-General Waves

Active Beneficiaries - In Table B.2 (below) the minimal detectable difference for the active beneficiary strata is a measure of the smallest difference between subgroups that 4,000 completes will be able to detect with 80 percent power and 90 percent confidence. For example, for a proportion of 0.10, a minimal detectable difference equal to 6.7 percentage points indicates that if 10 percent of the beneficiaries were employed who never attended college, and at least 16.7 percent of the beneficiaries were employed who attended at least

some college, the analysis will detect a significant difference between those who never attended college and those who attended at least some college. The table presents minimum detectable differences where one half of the sample is compared to the other half of the sample, and minimum detectable differences where 70 percent of the sample is compared to 30 percent.

Table B.2. Projected Minimal Detectable Differences Between Groups In Representative Beneficiary Sample

| Stratum | Half the Sample Compared to Other Half (2,000 vs. 2,000) | | | 70% of Sample Compared to 30% (2,800 vs. 1,200) | | |
|-----------------------|---|-------|-------|--|-------|-------|
| | Binomial Distribution | | | Binomial Distribution | | |
| | 10% | 30% | 50% | 10% | 30% | 50% |
| Overall (100 Percent) | 5.9% | 9.1% | 9.9% | 6.5% | 9.9% | 10.8% |
| Age 18 to 29 | 4.8% | 7.3% | 8.0% | 5.2% | 8.0% | 8.7% |
| Age 30 to 39 | 4.8% | 7.3% | 8.0% | 5.2% | 8.0% | 8.7% |
| Age 40 to 49 | 4.8% | 7.3% | 8.0% | 5.2% | 8.0% | 8.7% |
| Age 50 to 64 | 8.0% | 12.1% | 13.3% | 8.7% | 13.3% | 14.5% |

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

$$\text{Var}(\hat{p}) = \frac{p_1(1-p_1)}{n_{eff1}} + \frac{p_2(1-p_2)}{n_{eff2}}$$

where n_{eff1} and n_{eff2} are the effective sample sizes of the two populations being compared and $n_{eff1} = n_1 / deff_1$ and $n_{eff2} = n_2 / deff_2$. 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 = 0.10 and 80 percent power) are 2.49 square root ($\text{Var}(\hat{p})$).

Successful Workers - Of the 4,500 completed successful worker cases in round 2, 1,500 will be among SSI recipients who were working successfully as of the date six months prior to sample selection, 1,500 will be among DI beneficiaries in that period, and 1,500 will be among concurrent beneficiaries and recipients from both of these programs. Because we are interested in differences between successful workers and beneficiaries who are back on the rolls as of data collection and those who are not, a comparison of interest might be a comparison between these groups. Given that approximately 30 percent of successful workers could be back on the rolls at the time of data collection, such a comparison might involve a comparison between 70 percent of the sample (2,800 successful workers) and 30 percent of the sample (1,200

successful workers back on benefits). A comparison within strata (SSI, DI, or concurrent successful workers) would involve a comparison between 1,050 and 450 successful workers. Table B.3 (below) presents minimum detectable differences between the successful worker groups (sustained work vs. back on benefits), both for a comparison between two halves, and a comparison between 70 percent and 30 percent. These comparisons are also presented within strata (SSI, DI, and concurrent).

Table B.3. Projected Minimal Detectable Differences Between Successful Worker Groups

| Stratum | Half the Sample Compared to Other Half (2,250 vs. 2,250) | | | 70% of Sample Compared to 30% (3,150 vs. 1,350) | | |
|-----------------------|---|------|------|--|------|------|
| | 10% | 30% | 50% | 10% | 30% | 50% |
| Overall (100 Percent) | 3.5% | 5.3% | 5.8% | 3.8% | 5.8% | 6.4% |
| SSI only | 4.5% | 6.8% | 7.5% | 4.9% | 7.5% | 8.1% |
| SSDI only | 4.5% | 6.8% | 7.5% | 4.9% | 7.5% | 8.1% |
| Concurrent | 4.5% | 6.8% | 7.5% | 4.9% | 7.5% | 8.1% |

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

$$\text{Var}(\hat{p}) = \frac{p_1(1-p_1)}{n_{eff1}} + \frac{p_2(1-p_2)}{n_{eff2}}$$

where n_{eff1} and n_{eff2} are the effective sample sizes of the two populations being compared and $n_{eff1} = n_1 / deff_1$ and $n_{eff2} = n_2 / deff_2$. 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 = 0.10 and 80 percent power) are 2.49 square root ($\text{Var}(\hat{p})$).

b. Incentive Experiment

Due to declining response rates observed in the prior round 4 NBS, we plan to embed an experiment in the round 1 NBS-General Waves to determine whether an alternative incentive approach would maximize response at a lower overall cost. We will randomly assign two groups of beneficiaries to one of two experimental conditions (as seen in Table B.4 below)-both offering higher incentives for early responses. Sample members in the “early differential incentive” group would have the opportunity to earn a \$30 gift card if they complete the survey early in the interview period (for example, within the first two to four weeks) or \$20 for completing after that. Those in the “late differential incentive” group will have the opportunity to earn a \$30 gift card if they complete the survey in the two to

four weeks’ time period just prior to in-person interviewing attempts (approximately 12 weeks into the data collection period when a reminder letter is sent) or \$20 for completing after that. A third group, those receiving the standard \$20 gift card upon survey completion, will serve as the comparison group (with no offer of a higher incentive for early completion). We will target 450 completes in both experimental groups; the remaining 3,100 completed interviews will come from the comparison group.

Table B.4. Incentive Conditions

| Group | Completes | Incentive Condition |
|------------------------------|-----------|--|
| Early differential incentive | 450 | <ul style="list-style-type: none"> • \$30 gift card for call-ins during the first two weeks^a • \$20 gift card for remainder of data collection period |
| Late differential incentive | 450 | <ul style="list-style-type: none"> • \$20 gift card up until two weeks prior to CAPI • \$30 gift card for call-ins during two weeks preceding CAPI^a • \$20 gift card for remainder of data collection period |
| Standard incentive | 3100 | <ul style="list-style-type: none"> • \$20 gift card throughout data collection period |

^a We may consider lengthening the call-in period by an additional one to two weeks, if early call-in productivity warrants.

With 450 completes in each group, we could detect a 7.5 percentage point difference in completion rates between the two groups at 80 percent power. For example, a significant difference in the rate of completes within the first 2 weeks would be found if 17.5 percent of sample members in the differential group called in and completed the survey during that time period as compared to 10 percent of sample members in the standard incentive group

Cross tabulations will be developed to determine if there is a relationship between the incentive offering and timing and the rate of call-ins and the timeliness of response. An ANOVA will be used to determine if a significant difference exists between conditions.

B.2.5. Unusual Problems Requiring Specialized Sampling Procedures

NBS-General Waves - For the successful workers strata, there is a probability that the number of beneficiaries with successful work within the PSU areas, and within the three beneficiary types, will be too small to meet sample size requirements. If this occurs, as with the TTW participant sample in the prior NBS, we 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 CATI-only (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 NBS.

The result will be lower response rates for the non-PSU participants, and potential bias in the estimates. To address the bias issue, we will compare the responses of the within-PSU phone interview sample to those 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.

B.2.6. Periodic Cycles to Reduce Burden

We will administer the NBS-General Waves in 2015, 2017, and 2019. Sampled beneficiaries will complete the survey one time only (cross-sectional), with a new sample drawn prior to survey administration. Thus, there is no cyclic burden for respondents. However, a subset of beneficiaries selected for the successful worker strata will be followed longitudinally. We will follow successful workers who continue to be working at round 2 in round 3 so that we can better understand factors that positively or negatively impact the ability to sustain employment over time. To minimize burden, we will administer follow-up surveys biennially rather than annually, and will skip some items because they had previously been answered and are not prone to change. We will also use data from the previously completed survey to serve as question fills so that respondents’ cognitive burden of recalling previously reported employers and service providers is reduced.

The qualitative interviews will occur in 2015 only and thus pose no cyclic burden for participants.

B.3. Methods to Maximize Response Rates

B.3.1. Maximizing Response Rates

NBS-General Waves - Locating sampled beneficiaries and participants is our first challenge to obtaining a high response rate. While SSA has contact information for all potential respondents, we know from past experience that it will often not lead directly to the DI beneficiary or SSI recipient. 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 correspondence. These might, however, be a post office box, address of a guardian, financial institution, or other types of addresses that may make it difficult for us to locate the beneficiary or recipient. Since SSA now requires direct deposit of payment checks, we have minimized the importance of keeping address information current.

To improve contact information, we will mail an advance letter written on SSA letterhead and a study brochure to each sampled person prior to the survey, using the address of record (either from SSA administrative data or provider record). This letter describes the survey and indicates that we will soon contact the DI beneficiary or SSI recipient. In round 3, we will tailor the advance letter for longitudinal cases, as appropriate. We will begin locating with letters returned

to the contractor as undelivered. When an address is available without a phone number, we will conduct a directory search to obtain a number. When direct searches are unproductive, we will submit searches to Accurint, a comprehensive database compiled from multiple sources, and use locating letters, and telephone tracing (calling former neighbors or payees). In previous rounds of the NBS, we located more than 90 percent of DI beneficiaries and SSI recipients.

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

In the first three months of data collection, we will send locating letters, reminder letters, reminder postcards, and refusal conversion letters, as appropriate (see Attachment B, Respondent Correspondence). In round 3, we will tailor these for longitudinal cases, as appropriate. After two to three months of CATI interviewing, we will begin to transfer cases to field staff for locating. Delaying the start of field locating and interviewing allows an adequate number of cases to accumulate so field staff will have sufficient work and travel can be more cost effective. Prior to deploying field staff, we will send an advance letter to all sample persons with a valid mailing address, informing them that a representative from the contractor will be visiting their home (Attachment B). Once in the field, staff will have several other tools at their disposal to support field locating efforts, including a “Sorry I Missed You” card, appointment card, post-office letter, study brochure, interviewer field letter, and locating checklist. The locator checklist identifies steps a field interviewer should take when locating a respondent, with the steps listed hierarchically from most to least likely to be effective. The checklist helps prevent duplication of our efforts and sets clear parameters for when a case should cease because of lack of response.

The impairments and health of some DI beneficiaries and SSI recipients will make responding problematic, especially by telephone. To facilitate responses to the CATI interview, we will offer the use of several assistive devices (amplifier and 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 identify signs of respondent fatigue so that we can complete the interview 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.

To increase opportunities for self-response, we will permit assisted interviews, which differ from proxy interviews in that beneficiaries or recipients answer most questions themselves. The assistant, typically a family member, provides encouragement, interpretation, and verifies answers as needed. These interviews minimize item nonresponse, improve response accuracy, and help with some limiting conditions such as hearing difficulties and language barriers.

As a last resort, we will rely on proxy respondents to complete the survey on behalf of sample members who are unable to do so (even with assistance) either by telephone or in person. This includes individuals with severe communication impairments or physical disabilities that preclude participation in any mode, and those with mental impairments that might compromise data quality. The use of proxies can minimize the risk of nonresponse bias that would result from the exclusion of individuals with severe physical or cognitive impairments. To identify the

need for proxy respondents, we will administer a mini-cognitive test built into the prior NBS instrument. The test provides interviewers a tool for determining when to seek a proxy rather than leaving the decision to interviewer discretion or a gatekeeper. We will also develop and administer a Spanish-language of the instrument administered by Spanish-speaking interviewers to Spanish-speaking subjects. Translation interpretation services will be used for other non-English speakers.

The majority of respondents will receive a \$20 gift card to compensate them for their time, and we assure them of the confidentiality of their responses. Respondents assigned to one of two experimental groups examining the impact of a differential incentive on timeliness of survey completion will have the *opportunity* to receive a \$30 gift card if they call-in and complete the survey within a prescribed time period. We will provide a pre-paid gift card of \$5 in the final months of the data collection period to encourage call-ins from nonrespondents with the promise of a \$15 gift card after completion. These steps should also encourage sampled individuals to cooperate with the interviewer once contact is made.

To minimize burden for longitudinal respondents in round 3, we will administer follow-up surveys biennially rather than annually, and will skip some items because they had previously been answered and are not prone to change. We will also use data from the previously completed survey to serve as question fills so that respondents' cognitive burden of recalling previously reported employers and service providers is reduced.

Semi-Structured Interviews - To maximize participation among eligible beneficiaries and recipients with scheduled interviews, we will 1) schedule appointments at times convenient for beneficiaries and recipients (including evenings and weekends); 2) make reminder calls to them two business days prior to the scheduled interview; 3) reschedule appointments when asked to do so; 4) offer to break the interview into segments if beneficiaries or recipients' disability necessitates; 5) conduct interviews in Spanish, if appropriate; and 6) offer a \$20 gift card as a thank you for participation.

B.3.2. Dealing with Issues of Nonresponse

NBS-General Waves - We will adjust the base weights for survey nonresponse using the procedures described above and to control distributions for some variables to known totals from the administrative data. We can assess the extent of remaining bias by comparing weighted outcomes for the survey sample that can be observed in administrative data (for example, annual earnings and SSI and DI payments) to outcomes for the population that the weighted sample is intended to represent. We expect such comparisons to be especially important to assess attrition bias in analyses of the follow-up survey for the longitudinal samples. We will also be able to use the administrative data to assess the extent to which nonresponse in the follow-up survey is due to mortality.

Semi-Structured Interviews - There is a potential for non-response in the qualitative interview component of the study. This would occur when beneficiaries and recipients with scheduled interview appointments become unavailable for the interview. We will make all efforts to gain their cooperation, as detailed in the preceding sections. In the qualitative report, we will describe

the characteristics of non-responding participants so that the readers of the report can determine if the findings are relevant for their needs.

B.4. Tests of Procedures

NBS–General Waves - The original NBS survey items were developed and initially pre-tested as part of a separate contract held by Westat approximately ten years ago. 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 prior NBS contractor to prepare the instrument for CATI/CAPI programming, another pretest was conducted to ensure that the instrument was clear and understandable to respondents and to test interviewer usability. For NBS–General Waves, we removed two sections that were specific to the TTW program. Pretesting of the NBS-General Waves survey is not necessary, as no new items have changed or been added for round 1.

We anticipate making revisions to the survey instrument prior to round 2 to accommodate the addition of the successful workers. We will inform content through a review of the literature, a review of existing instruments, and, in large part, by the qualitative interviews with successful workers. After we make the revisions, we will test the revised instrument. We anticipate submitting the revised questionnaire as a request for non-material, non-substantive changes to OMB (i.e., change request).

Semi-Structured Interviews - In our preliminary work on this study, we conducted in-depth interviews with SSA beneficiaries to better understand factors that impacted their employment success, both positive and negative. SSA provided Mathematica with a sample of SSI, SSDI, and concurrent beneficiaries and recipients whose benefits and payments were suspended due to earnings. Recruiters contacted beneficiaries and recipients by telephone. During the initial recruitment call, we verified the sampled person’s name, address, and phone number, and determined whether the sample person was eligible for participation. We also offered a \$20 gift card as both an incentive to participate in the interviews, and a token of appreciation for their time. Our recruitment strategies and assumptions are based on this experience. We also based the interview guide for the proposed semi-structured interviews on these initial in-depth interviews.

B.5. Statistical Agency Contact for Statistical Information

The individuals consulted on technical and statistical issues related to data collection are listed in Table B.5.

Table B.5. Individuals Consulted on Technical and Statistical Issues

| Name | Affiliation/Address | Telephone Number |
|-----------------|---|--------------------|
| Elaine Gilby | Social Security Administration, Office of Research, Evaluation, and Statistics Washington, DC | (202) 358- 6449 |
| Paul O’Leary | | (202) 358- 6227 |
| Eric Grau | Mathematica Policy Research Washington, DC 20024 | (609) 945- 3330 |
| Gina Livermore | | (202) 264- 3462 |
| Frank Potter | | (609) 936- 2799 |
| David Stapleton | | (202) 484- 4224 |
| Debra Wright | | (202) 554- 7576 |
| Kirsten Barrett | | (202) 554- 7564 |

REFERENCES

Grau, Eric A., Kirsten Barrett, Debra L. Wright, Yuhong Zheng, Barbara Carlson, Frank Potter, and Sara Skidmore. "National Beneficiary Survey: Round 4 (Volume 1 of 3): Editing, Coding, Imputation, and Weighting Procedures." Washington, DC: Mathematica Policy Research, September 2011.

Mason, Mark. "Sample Size Saturation in PhD Studies Using Qualitative Interviews". *Forum: Qualitative Social Research*, vol. 11 no. 3, September 2010, Art 8.