1. Introduction

This appendix provides a discussion of survey methods used during the design and implementation of the Survey of Social Security Representative Payees and Beneficiaries (referred to as the Survey throughout this appendix). Section 1.1 of this chapter describes the background and objectives of the Survey. Section 1.2 provides an overview of the Survey and Section 1.3 describes the organization of the appendix.

1.1 Background of the Survey

The Survey was authorized by Section 1110 of the Social Security Act, as amended by Section 107 of Public Law 108-203 (Social Security Protection Act of 2004). That legislation required that the Commissioner of the Social Security Administration (SSA) conduct:

> "A statistically valid survey to determine how payments made to individuals, organizations, and State or local government agencies that are representative payees for benefits paid under Title II or XVI are being managed and used on behalf of the beneficiaries for whom such benefits are paid."

In addition, the legislation required that the Commissioner of SSA submit a report on the Survey to the Committee on Ways and Means of the House of Representatives and the Committee on Finance of the Senate.

The National Academies entered into a contract with SSA in October 2004 to study and make recommendations to SSA on how to improve the Representative Payee Program (RP Program). In turn, the National Academies contracted with Westat in October 2005 to design and implement the Survey. Specific tasks included designing the survey instruments and data collection protocol, selecting the survey sample, field testing and implementing the survey, processing the data, preparing the data sets and a data users

guide, and generating data tables for the National Academies to use to evaluate the RP Program.

SSA intends to use the Survey results to meet the legislative requirements mentioned above and to meet six evaluation objectives: (1) to assess the extent to which identified representative payees (RPs) are performing their duties in accordance with SSA standards for RP conduct; (2) to learn whether the representative payment policies are practical and appropriate; (3) to determine the extent to which the current program is serving the needs of beneficiaries with RPs; (4) to identify the types of RPs who have the highest risk of misuse of benefits; (5) to determine how well SSA supports RPs; and (6) to determine how knowledgeable RPs are about their roles and responsibilities.

1.2 Overview of the Survey

To meet the objectives, Westat conducted a cross-sectional probability sample survey of RPs and a survey of a subsample of beneficiaries who were linked to those RPs. The RPs were individuals who served fewer than 15 beneficiaries, and non-fee-for-service organizations that served fewer than 50 beneficiaries. A multistage stratified probability sample of RPs and their beneficiaries was selected from SSA administrative records. This entailed selecting a sample of 60 primary sampling units (PSUs); and, within PSUs, selecting a sample of 6,414 dyads (beneficiaries and associated RPs) from several different strata. For the subsample of beneficiaries who would be interviewed, one beneficiary (age 14 and older) was selected per RP for 50 percent of the RP sample. (See Chapter 2 for details of the sample selection.) The goal set for the Survey was to obtain completed interviews from 3,900 RPs and 1,950 beneficiaries. (Survey response rates are discussed in Chapter 4.)

The Survey used five survey instruments to collect data. Three instruments were administered to RPs (individual RP eligibility screener, organization RP eligibility screener, and RP interview) and two instruments were administered to beneficiaries (beneficiary capability screener and beneficiary interview). Instrument development entailed clarifying eligibility criteria and reflecting the criteria in the RP eligibility screener; modifying an existing validated instrument, the beneficiary capability screener, to measure beneficiary capability to participate in the Survey; and identifying and measuring the constructs implied by the evaluation objectives to be incorporated in the RP and

beneficiary interviews. Development also included cognitive testing and field testing of instruments. Once the instruments were finalized, they were translated into Spanish. Computer-assisted applications were then programmed to permit computer-assisted interviewing (CAI) by field interviewers.

Following approval of the Survey by the Office of Management and Budget (OMB), training and data collection began in July 2006. Preparation for data collection included prefield tracing of all organization RPs and only those individual RPs for whom new location information was obtained from electronic databases. Other predata collection activities included recruitment and training of field staff, and development and documentation of field procedures and training materials. (See Chapter 4 for a discussion of recruitment, training, and data collection activities.)

Survey instruments were administered to RPs and beneficiaries using CAI. In the case of the RPs, computer-assisted telephone interviewing (CATI) or computer-assisted personal interviewing (CAPI) was used. CAPI was used to collect data from beneficiaries (i.e., interviewers had to collect data from beneficiaries in person). In the event that a beneficiary was incapable of participating, CATI or CAPI was used to interview a proxy for the beneficiary.

On an ongoing basis, survey data was receipted, edited, and processed; and several quality assurance checks and procedures were implemented to ensure data security.[3] In addition, reports that were prepared pertaining to data collection progress and data quality were posted for the National Academies review on a secure website.

The sample design, instrumentation, and data collection permitted analysis at multiple levels. The RP survey, which gathered information on a specific selected beneficiary (referred to as the "reference beneficiary") and on the RP, permitted analysis at the level of beneficiaries who have RPs and at the level of RPs. The beneficiary survey, which gathered information on the reference beneficiary, permitted analysis at the level of beneficiaries age 14 and older who have RPs.

[3] All of the automated systems used for the Survey, including the CAI applications, adhered to a data security plan that was approved by SSA.

Once the data met quality standards, Westat statisticians weighted and developed variance estimates for the data. Three types of weights were developed, which correspond to the different levels of analysis mentioned in the preceding paragraph: (1) RP caseload weight, which ties to the universe of beneficiaries who have RPs; (2) RP level weight, which ties to the universe of RPs; and (3) beneficiary weight, which ties to the universe of beneficiaries age 14 and older who have RPs. For each type of weight, weighting entailed developing base weights, adjusted base weights, and poststratification adjustments. Separate replicate weights were created for the three types of RP and beneficiary weights. WesVar was used with the replicate weights and survey data to develop variance estimates that corresponded to the specific analyses conducted. (See Chapter 5 for a discussion of weighting and variance estimation.)

To prepare for analysis and delivery of final products from the Survey, staff constructed several data files, with each file including the survey data, composite variables, and the weights. A data user's manual was created, which documented the data files.

Using the final data files, Westat analysts conducted analyses that tied to the evaluation objectives. These analyses were primarily descriptive. As appropriate, inferential tests were performed pertaining to the stated hypotheses. The results of the analyses supported the narrative report that the National Academies prepared for SSA.

1.3 Organization of Appendix

The appendix is organized in seven chapters. Chapter 1 provides background and an overview of the Survey. Chapter 2 contains information about the sample design and implementation, i.e., definition of the target population, selection of primary and secondary sampling units, selection of dyads and beneficiaries, and overall selection probabilities. Chapter 3 provides information on survey instrument development and testing, including cognitive testing and field testing, translation of instruments, and development of CAI applications. Chapter 4 describes the data collection activities for the Survey. Topics include field staff recruitment and training, procedures used during data collection, data security, data editing and cleaning procedures, quality control efforts connected with data collection, and calculation of RP and beneficiary response rates. Chapter 5 contains information on the development of analytic weights (i.e., base weights, nonresponse adjustment of weights, post stratification (raking) adjustments, and a

description of the nonresponse bias study) and information about the calculation of variance estimates (i.e., replicate weights, confidence intervals, and design effects). Chapter 6 summarizes the documentation of the Survey data files and the quality assurance files, and describes the procedures for destroying files and materials after final data delivery. Chapter 7 contains information on analysis activities. Topics include, development of the analysis plan, recoding of variables and creation of aggregate variables, development of frequency distributions and bivariate analyses, examination of proxy responses, hypothesis testing, table production, and quality control of tabulations.

2. Sample Selection

This chapter discusses the selection of the representative payee (RP) sample and the beneficiary sample. This includes a description of the sample design and summary of sample design analyses that were performed prior to sample selection.

2.1 Target Populations

The Survey of Social Security Representative Payees and Beneficiaries (Survey) has two target populations: one of RPs and the other of beneficiaries. The source information for membership in the target populations was Social Security Administration (SSA) data, including the Representative Payee System (RPS) updated in January 2006, and finder files created from other SSA files in March 2006. The RP target population was restricted to individual RPs serving fewer than 15 beneficiaries and non-fee-for-service organization RPs serving fewer than 50 beneficiaries. The target RP target population was further restricted to RPs satisfying all of the following conditions:

- Condition 1: Residing inside the 48 contiguous states;
- Condition 2: Serving as an RP to someone other than oneself;
- Condition 3: Having one or more current beneficiaries;
- Condition 4: Having a valid state/county code; and
- Condition 5: Managing funds of more than $50 each month for one or more beneficiaries.

Administrative data in SSA's RPS was used to exclude ineligible RPs from the target RP population on the basis of eligibility conditions 1 through 4. The testing of eligibility Condition 5, however, required issuance data not present in the RPS. This condition could be tested using data from other SSA files, but the number of RPs for which issuance data could be obtained was restricted by SSA to less than 10,000.

The target population of beneficiaries was restricted to all persons age 14 and older whose benefits were more than $50 each month and were managed by an eligible RP.

2.2 Overview of the Sample Design

This section summarizes the Survey objectives, sampling methods, sample sizes, and notation used in the remainder of the chapter.

2.2.1 Survey Objectives

The major analysis goals for the Survey involved comparing estimated proportions among the following subpopulations:

- Compare between types of payees:
  - Organization RPs; and
  - Individual payees.
- Compare among beneficiary-age categories:
  - Children;
  - 18 to 64; and
  - 65+.
- Compare between programs:
  - SSI; and
  - OASDI.
- For individual RPs, compare among relationships of RP to beneficiary:
  - Parent;
  - Adult child or other relative;
  - Unrelated RP who has only one beneficiary; and
  - Unrelated RP who has two or more beneficiaries.
- For individual RPs, compare between living relationships:
  - Living with beneficiary; and
  - Not living with beneficiary.

2.2.2 Sampling Methods

Multistage sampling was used to select two samples for purposes of interviewing RPs and beneficiaries. In terms of the sampling units, one was a sample of *dyads*, where a dyad is a beneficiary and the associated RP; and the second, which was also a subsample of the first, was a sample of beneficiaries. The ultimate sampling units for the two samples, however, were not the reporting units for these two samples. In the sample in which the beneficiary was the ultimate sampling unit, data were collected from both RPs and beneficiaries. In the portion of the sample in which dyads were the ultimate sampling units and the beneficiaries in these dyads were not selected for the subsample of beneficiaries, data were collected from only the RPs. In the remainder of this discussion about sample selection, the selected samples are referred to with respect to their ultimate sampling units, not with respect to their reporting units.

The stages of sampling were the following:

- The first stage was a national sample of clusters of counties, called *primary sampling units* (PSUs). The PSU sampling frame was stratified by demographic characteristics and regional location, and one PSU was selected from each stratum using probability-proportional-to-size (PPS) sampling, in which the PSU measure of size (MOS) was total population according to the 2000 Decennial Census.

- The second stage was a sample of *secondary sampling units* (SSUs), which were clusters of ZIP codes within the sampled PSUs. The SSUs were created using a minimum-MOS criterion and were selected using systematic PPS sampling. SSU-level counts of the number of dyads within analysis domains (described below) were used to compute the MOS for SSUs.

- The third stage was a stratified two-phase sample of dyads. The conditional probabilities of selecting dyads were equal within the dyad sampling strata, resulting in RPs being selected with probabilities proportional to the number of beneficiaries they represented. The first phase sample was selected from dyads with RPs satisfying eligibility conditions 1 through 4. The first-phase sample was then screened by determining eligibility under Condition 5 and then subsampled, yielding a second phase sample that contained no ineligible RPs on the basis of eligibility conditions 1 through 5. The RP of each dyad selected at this stage was a sampled RP. The beneficiary of each dyad selected at this stage was the *reference beneficiary*, who was referred to by name in the questionnaire administered to the sampled RP. Because the data collection

protocol did not allow for an RP to be included in the dyad sample more than once with different reference beneficiaries, sets of sampled dyads with the same RP were subsampled by selecting with equal probability one dyad from each such set.

- The fourth stage excluded dyads in which the beneficiary was younger than 14 and subsampled those dyads in which the age of the beneficiary was greater than or equal to 14. The beneficiary in each dyad selected at this stage was included in the beneficiary sample.

2.2.3 Sample Sizes

Some of the subpopulations described in Section 2.2.1 are very small compared to the entire population of dyads. Consequently, it was necessary to stratify the dyad population and then oversample some dyad strata. Table 2-1 contains the definitions for the dyad sampling strata, which were intersections of subpopulations of interest and were not individually of interest as analysis domains. The 13 domains listed in column 2 of Table 2-2 were used to control the allocation of the RP and beneficiary samples to the 27 dyad strata so that, except for the organization-RP domain, the effective sample sizes for each domain were approximately equal. For the organization-RP domain, the effective sample size was one-half of the effective sample size for each of the other 12 domains. The allocation of the RP and beneficiary samples to the dyad strata used the multivariate allocation procedure described by Chromy (1987) and software described by Zayatz and Sigman (1995). In addition to determining sample sizes by dyad stratum and domain (see columns 3 and 4 of Table 2-2), the allocation of the samples determined for each dyad stratum used a *disproportionality factor*, which was the ratio of the stratum's proportion of the dyad sample to its proportion of the dyad population. Factors greater than 1.0 indicated relative oversampling, and factors less than 1.0 indicated relative undersampling. Across the 27 dyad strata, the disproportionality factors ranged from 0.3 to 14.3. (See column 2 of Table 2-3.)

Table 2-1. Dyad stratum definitions

Stratum number	Stratum definition			
	Type of RP	Beneficiary age[1]	RP Relationship[2]	Living arrangement
1	Individual	18 to 64	Unrelated-2+	Not with
2	Individual	18 to 64	Unrelated-2+	With/Unknown[3]
3	Individual	18 to 64	Unrelated-1	Not with
4	Individual	18 to 64	Unrelated-1	With/Unknown
5	Individual	18 to 64	Adult Child/Other Relative	Not with
6	Individual	18 to 64	Adult Child/Other Relative	With/Unknown
7	Individual	18 to 64	Parent	Not with
8	Individual	18 to 64	Parent	With/Unknown
9	Individual	18 to 64	Unrelated-2+	Not with
10	Individual	65+	Unrelated-2+	With/Unknown
11	Individual	65+	Unrelated-1	Not with
12	Individual	65+	Unrelated-1	With/Unknown
13	Individual	65+	Adult Child/Other Relative	Not with
14	Individual	65+	Adult Child/Other Relative	With/Unknown
15	Individual	65+	Parent	Not with
16	Individual	65+	Parent	With/Unknown
17	Individual	Child	Unrelated-2+	Not with
18	Individual	Child	Unrelated-2+	With/Unknown
19	Individual	Child	Unrelated-1	Not with
20	Individual	Child	Unrelated-1	With/Unknown
21	Individual	Child	Adult Child/Other Relative	Not with
22	Individual	Child	Adult Child/Other Relative	With/Unknown
23	Individual	Child	Parent	Not with
24	Individual	Child	Parent	With/Unknown
25	Organization	Child		
26	Organization	65+		
27	Organization	18 to 64		

[1] For the RP population, "child" is ages 0 through 17; for the beneficiary population, ages 14 through 17.

[2] "Unrelated-1" denotes an RP who has only one unrelated beneficiary. "Unrelated-2+" denotes an RP who has two or more unrelated beneficiaries.

[3] Unknown living arrangement grouped with living-with living arrangement in order to increase the probability of selecting dyads having not-living-with living arrangements.

Table 2-2. Analysis domains and associated number of fielded RP and beneficiary interviews

Domain number	Description	Fielded number of RP interviews	Fielded number of beneficiary interviews
1	RPs representing beneficiaries with age 18-64	2,601	1,446
2	RPs representing beneficiaries with age 65+	1,032	578
3	RPs representing beneficiaries who are children	1,465	519
4	Individual RPs representing two or more unrelated beneficiaries	858	451
5	Individual RPs representing only one unrelated beneficiary	884	438
6	Individual RPs who are either an adult child or another relative of the beneficiary they represent	1,390	716
7	Individual RPs who are a parent of the beneficiary they represent	1,501	701
8	Individual RPs who are living with the beneficiary they represent or the living arrangement is unknown	3,622	1,775
9	Individual RPs who are not living with the beneficiary	1,011	531
10	RPs representing a beneficiary receiving SSI benefits	2,485	786
11	RPs representing a beneficiary receiving OASDI benefits	3,133	1,618
12	Individual RPs	4,633	2,306
13	Organizational RPs	465	237

Table 2-3. Disproportionality factors and average sampling rates by dyad stratum

Stratum number	Dyad-sample disproportionality factor	Dyad-sample average sampling rate; 1 out of:	Beneficiary-sample average sampling rate; 1 out of:
1	14.2	80.2	144.0
2	13.7	69.6	124.8
3	8.1	118.1	210.2
4	7.2	139.7	250.8
5	3.8	256.4	460.4
6	0.6	1773.2	3182.7
7	3.8	259.6	464.8
8	0.4	2000.4	3585.0
9	14.3	99.5	171.8
10	13.8	65.9	117.2
11	8.3	125.7	232.1
12	7.4	143.6	255.0
13	4.3	207.7	372.2
14	2.0	507.6	911.3
15	4.3	224.7	449.4
16	2.0	405.5	811.0
17	14.2	71.0	147.8
18	13.7	83.1	134.3
19	8.1	68.9	69.0
20	7.2	143.0	255.9
21	3.8	208.0	378.1
22	0.6	1765.2	3158.2
23	3.8	234.5	420.0
24	0.3	3137.7	3159.7
25	3.9	31.4	50.5
26	4.4	262.4	471.4
27	3.9	237.3	425.5

2.2.4 Notation

In the sections that follow, the probabilities of selection are specified for each stage of sampling, using the following notation:

$P(\alpha)$ = probability of selecting PSU α,

$P(\beta \mid \alpha)$ = conditional probability of selecting SSU β in PSU α,

$P(\delta \mid \alpha\beta)$ = conditional probability of selecting dyad δ in SSU β in PSU α, and

$P(\varepsilon \mid \alpha\beta\delta)$ = conditional probability of selecting beneficiary ε in dyad δ in SSU β in PSU α.

2.3 PSU Sample

The sampling frame for the first stage of sampling consisted of PSUs that were either single counties or groups of contiguous counties, depending on population size. The sampling frame was stratified into 60 strata based on the PSU characteristics of region, MSA-versus-non-MSA, per capita income, percent Hispanic, percent non-Hispanic African Americans, and percent nonminority. Each of two strata contained only one PSU, which was selected with certainty into the PSU sample. The other 58 strata were noncertainty strata, from each of which one PSU was selected with probability proportional to 2000 Census population. The remainder of this section provides additional details about the PSU sampling frame and the selection of the PSU sample.

The PSU sampling frame was a modified version of a frame of 1,884 PSUs constructed from the 3,141 U.S. counties in 2002 for the Adult Literacy and Lifeskills (ALL) Survey. The PSUs in the ALL Survey frame and in the modified frame were identical except that 16 PSUs in Alaska and Hawaii were dropped from the modified frame because the current survey was confined to the contiguous United States.

The number of PSUs to be sampled from the frame for the current survey was set at 60, which happened to the same number as that sampled for the ALL Survey. Consideration was given to the possibility of using the PSUs sampled for the ALL Survey to be used for the Survey of Representative Payees and Beneficiaries. Doing so would yield considerable fieldwork benefits. As a result of conducting the ALL survey in 2002, Westat

had the names and addresses of over 300 experienced field interviewers who had received high performance ratings and were located in ALL-survey PSUs. Consequently, costs to recruit interviewers would have been higher if new PSUs were selected for the current survey. Also, because the proportion of interviewers for which replacements must be obtained during the survey cycle is higher for newly hired interviewers than it is for experienced interviewers, interviewer training costs were expected to be higher if a fresh sample of PSUs were selected. However, some loss in sampling efficiency would result from using the ALL Survey PSU sample compared with using a sample design tailor-made for the Survey of Representative Payees and Beneficiaries.

A comparison was made between the predicted precision resulting from using the ALL Survey PSUs selected with probability proportional to 2000 Census population with those from reselecting the PSU sample using a composite MOS described in Folsom, Potter, and Williams (1987). The composite MOS was calculated from PSU-by-dyad-stratum counts of RPs and the disproportionality factors for the dyad strata. Overall, the gains in precision from reselecting an independent PSU sample for the current survey were judged to not outweigh the data collection benefits resulting from using the ALL sampled PSUs for the current survey.

To control sampling variance, having each stratum's total MOS be approximately the same was desirable. Following the exclusion of PSUs in Alaska and Hawaii, a second modification of the frame was to rebalance the stratum total MOS in two strata. One PSU in a stratum with an above average total MOS was reassigned to a similar stratum with a much reduced total MOS, resulting from discarding PSUs in Alaska and Hawaii. The net effect of these two modifications to the ALL sampling frame was that the total stratum MOS changed in five strata. In four of these five strata, total stratum MOS decreased because of the discarding of PSUs in Alaska and Hawaii. In one stratum, total stratum MOS increased because its gain in MOS from receiving the PSU transferred from another stratum was greater than its loss in MOS from the discarding of PSUs in Alaska and Hawaii.

Since PPS sampling was being used to select the PSU sample, the selection probability for a PSU was equal to its MOS divided by its total MOS for the PSU stratum to which it had been assigned. Hence, in the 54 PSU strata in which no change occurred in total stratum MOS, no changes in PSU selection probabilities resulted. In the five PSUs in which the total stratum MOS decreased, the PSU selection probabilities increased. In the

one PSU that total stratum MOS increased, the PSU selection probabilities decreased. Kish and Scott (1971) describe several procedures for retaining PSUs in a sample when their selection probabilities change. The procedure that they refer to as Procedure B applies to the PSU sample for the current survey. In accordance with Procedure B, the 59 sample PSUs in the ALL survey for which no change in selection probability occurred or an increase in selection probability occurred were retained for use in the current survey. In the one PSU stratum in which the PSU selection probabilities decreased, a new PSU was selected. This new sample PSU replaced a sample PSU that had been discarded from the stratum because it was outside of the contiguous United States.

2.4 SSU Sample

To further control data collection costs in a few large sampled PSUs, smaller geographic units were selected in a second stage of sampling. SSUs were created by first clustering dyads within PSUs by ZIP code. These clusters were called *ZIP fragments* because each cluster corresponded to the portion of a ZIP code contained within a PSU. The following composite MOS was calculated for each ZIP fragment:

$$mos_{\alpha f} = \sum g_h M_{\alpha fh},$$

where g_h is the disproportionality factor for dyad-stratum h and $M_{\alpha fh}$ is the number of dyads assigned to dyad-stratum h in ZIP fragment f in PSU α. The ZIP fragments were sorted by PSU and ZIP code, and adjacent ZIP fragments within the same PSU were then combined to create SSUs such that the aggregate MOS for the SSU was not less than

$$MINMOS = 8\max[14(\max_{1\le h\le 24} g_h), 49(\max_{25\le h\le 27} g_h)].$$

The constant 8 appears in the formula for *MINMOS* because it was desired to assign at least eight RP interviews to an SSU, so that approximately four or more in-person interviews with beneficiaries could be conducted in an SSU. The expression involving the disproportionality factors appears in the formula for *MINMOS* so that in PSUs containing SSUs the sampling interval for the sampling of dyads would be large enough to prevent two or more dyads with the same RP being included in the dyad sample. (In strata 1 through

24, up to 14 dyads could have the same RP; and in strata 25 through 27, up to 49 dyads could have the same RP.)

The target number of sampled SSUs in a PSU was the expected number of RP interviews in the PSU divided by eight. If this number was greater than the number of SSUs created in a PSU, the sampling of SSUs from the PSU was abandoned and the subsequent sampling of dyads was from the entire PSU, not from SSUs. The sampling of SSUs occurred in 16 PSUs. The number of SSUs created in these PSUs ranged between 12 and 66. These SSUs were sorted by PSU and ZIP code, and systematic stratified PPS sampling of the SSUs was performed using the following MOS:

$$mos_{\alpha\beta}^{(SSU)} = \sum_h g_h M_{\alpha\beta h} .$$

The number of sampled SSUs in PSU α, denoted n_α, was approximately equal to the expected number of fielded RP interviews in the PSU divided by 8. In the two certainty PSUs, an even number of SSUs was selected to facilitate variance estimation method during data analysis. The number of sampled SSUs per PSU ranged between 8 and 20. For PSUs containing SSUs, the SSU selection probability for PSU β was

$$P(\beta \mid a) = n_\alpha mos_{\alpha\beta}^{(SSU)} / MOS_\alpha^{(SSU)} \approx \frac{(5100) mos_{\alpha\beta}^{(SSU)} / P(\alpha)}{8 \sum_{\alpha'} MOS_{\alpha'}^{(SSU)} / P(\alpha')},$$

where $MOS_\alpha^{(SSU)}$ is the sum of $mos_{\alpha\beta}^{(SSU)}$ over all SSUs (both sampled and unsampled) in PSU α.

2.5 Sampling of Dyads

The target number of completed RP interviews for the dyad sample was 3,900. January 2006 RPS data was used to allocate the target sample size to the 27 dyad strata by controlling the decreases in effective sample size resulting from disproportionate allocation.

A dyad sampling frame was constructed, consisting of all dyads in the RPS as of late November 2005, for which the most recent RP address was within a sampled SSU area (excluding those not satisfying one or more of eligibility conditions 1 through 4). The total size of the first-phase sample was 6,414, the result of increasing the target of 3,900

completed RP interviews (and the associated dyad-stratum allocations). This allowed for nonresponse of up to 20 percent, for field ineligibility of up to 5 percent, and for exclusion of up to 20 percent of the first-phase sample due to a failure to satisfy eligibility Condition 5. The sampling frame was sorted by dyad strata. Within each stratum, the dyads were listed by PSU, SSU within PSU, and RP. The first-phase sample of dyads was a systematic PPS sample selected from this ordered list.

The MOS for the stratified PPS sampling of dyads was

$$mos_{\alpha\beta\delta}^{(dyad)} = 1/[P(\beta \mid \alpha)P(\alpha)].$$

If a dyad's MOS was larger than the MOS sampling interval, the dyad was selected with certainty—that is, $P(\delta \mid \alpha\beta) = 1$. Dyads selected with certainty, which were 4.4 percent of the total dyad sample, were those with large positive disproportionality factors (i.e., that were in small analysis domains) belonging to low-population PSUs (that had been not subdivided into SSUs). Following the selection of the certainty dyads, noncertainty dyads were selected with probabilities

$$P(\delta \mid \alpha\beta) = (m_h - m_h^*)(mos_{\alpha\beta\delta}^{(dyad)})/(MOS_h^{(dyad)} - MOS*_h^{(dyad)}),$$

where m_h and m_h^* are the number of dyads and the number of certainty dyads, respectively, selected from stratum *h*; and $MOS_h^{(dyad)}$ and $MOS*_h^{(dyad)}$ are the stratum totals of $mos^{(dyad)}$ for all dyads and all certainty dyads, respectively, in stratum *h*.

Within each dyad stratum, the sampling frame was sorted so that dyads with the same RP were next to each other. This minimized the chance that two or more dyads with the same RP would be selected. Nevertheless, 514 of the 6,414 sampled dyads had the same RP as another sampled dyad. There were 180 sets of sampled dyads with the same RP: 115 of these sets contained 2 sampled dyads with the same RP, 47 sets contained 3 or 4 sampled dyads with the same RP, and the remaining 18 sets contained between 5 and 11 sampled dyads with the same RP. Sets of sampled dyads that had the same RP were subsampled by selecting with equal probability one dyad from each set. Subsampling of dyads with the same RP reduced the size of the first-phase sample of dyads to 6,080 dyads, producing unconditional first-phase dyad-selection probabilities of

$$P(\alpha\beta\delta^{(1)}) = P(\alpha\beta\delta)/d,$$

where *d* was the number of sampled dyads having the same RP as the subsampled dyad.

SSA provided the issuance data for each beneficiary in a sampled dyad in the first-phase sample. Based on these administrative data, 888 first-phase sampled dyads were classified as ineligible. Table 2-4 describes reasons for ineligibility based on administrative data and the number of sample dyads designated as ineligible for each of these reasons.

Table 2-4. Reasons for ineligibility based on administrative data

Reason	Number of ineligible dyads
The amount of the monthly issuance to the reference beneficiary was $50 or less.	811
Unable to match monthly issuance data to the reference beneficiary's SSA identifier in the RPS.	42
The RPS and the issuance data disagreed on the name of the reference beneficiary.	33
Two sampled RPs had the same beneficiary. The dyad with the older effective date for the RP address was designated as ineligible.	2

According to SSA, current RP practice is that a beneficiary can have only one RP. Prior to the adoption of this practice, however, a beneficiary that received both SSI and OASDI could have a different RP for each program. Dyads designated as ineligible because two sampled RPs had the same beneficiary had not had their RPS information updated to reflect current RP practice, or the dyads were not following the current RP practice.

The dyad-stratum allocation for the second-phase sample of dyads was determined by increasing the target number of completed interviews in each dyad stratum to allow for up to 20 percent nonresponse and up to 5 percent field-determined ineligibility. Within each dyad stratum, a second-phase sample of dyads was selected from eligible first-phase sampled dyads with probability proportional to $1/P(\alpha\beta\delta^{(1)})$. This method of selecting the second-phase sample produced equal unconditional second-phase dyad-selection probabilities,

$$P(\alpha\beta\delta^{(2)}) = P(\delta^{(2)}|\alpha\beta\delta^{(1)})\,P(\alpha\beta\delta^{(1)}),$$

within a dyad stratum for all second-phase sampled dyads not selected with certainty from eligible first-phase sampled dyads—that is, for second-phase sampled dyads with $P(\delta^{(2)}\,|\,\alpha\beta\delta^{(1)}) < 1$. The actual

number of dyads in the second-phase sample was 5,105, slightly less than 5,130. Across all 27 dyad strata, 300 dyads (equal to 6% of the resulting second-phase sample size) were selected in the first-phase sample and determined to be eligible, but were not selected for the second-phase sample.

Following the selection of the second-phase dyad sample, addresses and telephone numbers for the sampled RPs and beneficiaries were obtained from an address and a telephone vendor. When the information obtained from the vendors disagreed with information in the RPS, tracing operations were performed by telephone tracers to obtain the addresses and telephone number to be used for data collection. The tracing operation determined that, for seven pairs of sampled dyads, the RPs for the dyads in each pair had different names in the RPS, but they were actually the same RP associated with different sets of beneficiaries. One of the dyads in each of these pairs was deleted and the undeleted dyad was assigned a conditional selection probability of 0.5. Following the deletion of seven dyads that had been determined to have the same RPs as other sampled dyads, 5,098 sampled dyads were assigned to data collection for the interviewing of each dyad's RP. The third column of Table 2-2 tabulates the fielded number of RP interviews by various analysis domains, based on RPS values for the sampled dyad's domain-membership variables.

2.6 Sampling of Beneficiaries

The target number of completed beneficiary interviews was half that for RPs, i.e., 1,950. Consequently, the desired size of the beneficiary sample was set at 2,565, the result of increasing the target number of 1,950 completed beneficiary interviews to allow for up to 20 percent nonresponse and for up to 5 percent field ineligibility. The beneficiary sample was selected as a subsample of the reference beneficiaries from the dyad sample, after excluding beneficiaries younger than 14. No further subsampling was done for beneficiaries age 14 to 17 in dyad stratum 24—in which the RP is a parent living with the beneficiary—because of the large number of beneficiaries removed from this stratum by excluding beneficiaries younger than 14. (The dyad strata were the result of intersecting desired domains for analyzing RP data. Consequently, no dyad stratum contained only sampled dyads with beneficiaries age 14 to 17, which would provide more control over the number of beneficiaries age 14 to 17 in the beneficiary sample.) All the other dyad strata were subsampled by sampling across these strata with selection probabilities proportional to $g_h / P(\alpha\beta\delta^{(2)})$, in order to achieve the desired sample size for the beneficiary sample**.** This produced equal unconditional beneficiary-selection probabilities,

$$P(\alpha\beta\delta^{(2)}\varepsilon) = P(\varepsilon|\alpha\beta\delta^{(2)})\,P(\alpha\beta\delta^{(2)}),$$

within a dyad stratum for beneficiaries not subsampled with certainty—that is, for subsampled beneficiaries with $P(\varepsilon \mid \alpha\beta\delta^{(2)}) < 1$.

The actual number of eligible beneficiaries in the beneficiary sample was 2,549. As mentioned earlier, tracers determined that 7 dyads had the same RPs. One dyad was deleted from each of 6 of those dyads where both RPs and beneficiaries had been selected for interview, resulting in the deletion of 6 sampled beneficiaries. Thus, 2,543 sampled beneficiaries were assigned to data collection for the interviewing of beneficiaries. The fourth column of Table 2-2 tabulates the fielded number of beneficiary interviews by various analysis domains, based on RPS values for the sampled beneficiary's domain-membership variables.

2.7 Overall Selection Probabilities

For use in the calculation of sampling weights, the following unconditional selection probabilities were determined for each sampled dyad and beneficiary:

$P(\alpha\beta\delta)$ = unconditional probability of selecting dyad δ in SSU β in PSU α

$P(\alpha\beta\delta\varepsilon)$ = unconditional probability of selecting beneficiary ε in dyad δ in SSU β in PSU α

These unconditional probabilities were calculated by using the following equations:

$$P(\alpha\beta\delta) = P(\alpha)P(\beta \mid \alpha)P(\delta \mid \alpha\beta)$$

and

$$P(\alpha\beta\delta\varepsilon) = P(\alpha\beta\delta)\;P(\varepsilon|\alpha\beta\delta)\,,$$

where the conditional probabilities on the right-hand sides of these equations are defined above.

For each sampled RP or beneficiary, the reciprocal of the unconditional selection probability is the associated overall sampling rate. For example, if an unconditional selection probability is 0.02, the associated sampling rate is 1 out of 50. Columns 4 and 5 of Table 2-3 contain the average sampling rates by dyad strata for the RP and beneficiary sample, respectively.

References

Chromy, J. (1987). Design Optimization with Multiple Objectives. *Proceedings of the Survey Research Methods Section*. American Statistical Association. 194-199.

Folsom, R., Potter, F., and Williams, S. (1987). Notes on a composite size measure for self-weighting samples in multiple domains. *Proceedings of the Survey Research Methods Section,* American Statistical Association. 250-255.

Kish, L. and Scott, A. (1971). Retaining units after changing strata and probabilities. *Journal of the American Statistical Association.* 66, 461-470.

Zayatz, L. and Sigman, R. (June 1995). CHROMY-GEN: General-Purpose Program for Multivariate Allocation of Stratified Samples Using Chromy's Algorithm. *Economic Statistical Methods Report ESM-9502*. Bureau of the Census, Washington, DC.

3. Survey Instrument Design and Testing

This chapter contains a discussion about survey instrument design and testing, including cognitive testing and field testing, translation of instruments, and development of Computer Assisted Interviewing (CAI) applications.

3.1 Development of Instruments

Instruments for the Survey of Social Security Representative and Beneficiaries (Survey) included an eligibility screener administered to representative payees (RPs); a capability screener administered to beneficiaries; and interview instruments administered to both RPs, beneficiaries, and when necessary to proxies for beneficiaries. All instruments were administered using CAI. The RP eligibility screener was used to assess the eligibility of both RPs and beneficiaries, and to provide input for the RP and beneficiary interviews to determine skip patterns. Two versions of the RP eligibility screener were used: the individual RP screener and the organization RP screener. The beneficiary capability screener was used to determine the beneficiary's ability to understand and provide consent to participate in the Survey. Eligible beneficiaries who passed the capability screener were asked to complete the beneficiary interview. Screeners were developed based on screeners previously used by Westat on other surveys.

For the development of the instruments, the Social Security Administration (SSA) provided a draft list and a description of topics for each instrument and a crosswalk between the topics and the evaluation objectives. In collaboration with the National Academies and the National Academies Committee on Social Security Representative Payees (Committee), Westat created questions to measure the topics, formatted instruments, and clarified and added or subtracted questions after testing the instruments. The Committee, NA, and Westat worked closely during the questionnaire development phase to make sure that the survey collected the data needed to meet the evaluation objectives as stated by SSA (see Section 1.1). Doing this required NA to collaborate and clarify goals and procedures with SSA and to interface with others involved in the RP system. In addition, items were added that asked how the RP came to take on this role for this beneficiary – who made the request, their willingness to serve, and their satisfaction

with their performance. Furthermore, design staff laid the groundwork for the creation of aggregate variables or scales within each topic area by using consistent time references, similar response categories, and by including questions of similar content and wording among the instruments, as appropriate. Cognitive testing of the interview instruments and field testing of the screeners and the interview instruments provided useful information for revising the survey instruments.

3.1.1 Computer Assisted Interviewing

CAI was used to administer the instruments, as it offered a number of benefits. CAI allows complex skip patterns to be administered smoothly. It is especially helpful when skips incorporate information gathered earlier in the interview or from administrative data. CAI also allows for automated filling of information such as the RP's or beneficiary's name and the beneficiary's age. In addition, it incorporates real time error checks of values and questions. Finally, CAI allows for rapid transfer of interview information to database, facilitating interviewer supervision and rapid feedback, as well as faster data cleaning and analysis.

3.1.2 Instrument Flow and Pathways

The RP and beneficiary interview instruments were designed to flow from more general to more specific questions. More sensitive questions were placed at the end of an interview because (1) rapport was likely to be higher after spending time with the interviewer and (2) it reduced risk that refusal from the more sensitive questions would lead to discontinuation of the interview earlier.

Skip patterns were built into the screeners and interview instruments. All screeners included skips based on whether the respondent reported having read the introductory letter and the Privacy Act statement at the bottom of the letter, and whether they had questions related to these. Other skip patterns were particular to some screeners but not others.

- The beneficiary capability screener was administered if the beneficiary lived in a supervised setting, was an adult with a legal guardian, or if the

RP or interviewer had concerns about the beneficiary's ability to complete the questionnaire.

- Both the individual and organization RP screeners incorporated skip patterns based on whether the beneficiary had been selected for participation.

- The organization RP screener incorporated an opportunity to change respondents if the interviewer determined that the initial respondent was not "the most senior knowledgeable person in [the] organization who knows and takes care of [the] beneficiary's needs and who is familiar with how [the] beneficiary's Social Security payments are used."

Many of the interview skip patterns were based on data collected during the RP eligibility screening. Skip patterns for the beneficiary interview instrument were based on whether the beneficiary was a minor, had only one RP in his or her lifetime, lived with his or her RP, whether the RP was the beneficiary's parent, the type of benefit received (e.g., Title II or Title XV1), and whether the beneficiary was institutionalized.

The RP interview instrument incorporated skip patterns based on whether the RP was an individual or an organization, whether the respondent was an organization RP administrator, the beneficiary's age, the beneficiary's ability to communicate, where the beneficiary lived, and the type of benefit received.

The administration time goals of the survey instruments were about 5 minutes for the RP and beneficiary screeners and about 30 and 20 minutes, respectively, for the RP and beneficiary interview instruments. However, the field test demonstrated that administration of the survey instruments was more time consuming than expected. After the field test, screeners and interview instruments were modified to bring administration time more in line with the timing goal.

3.1.3 Instrument Content

Instruments for the Survey of Social Security Representative and Beneficiaries (Survey) included an eligibility screener administered to representative payees (RPs); a capability screener administered to beneficiaries; and interview instruments administered to RPs, beneficiaries, and proxies for beneficiaries when necessary. All instruments were

administered using CAI. Hard copy versions of the screeners and questionnaires are in Appendix B.

RP Eligibility Screeners

The RP eligibility screener was used to assess the eligibility of both RPs and beneficiaries, and to provide input for the RP and beneficiary interview instruments to determine skip patterns. There were two versions of the RP eligibility screener: the individual RP eligibility screener and the organization RP eligibility screener.

Because the RPs' or beneficiaries' circumstances could change over time, the screening process was designed to confirm the administration data received from the Social Security Administration (SSA) by Westat, and to gather updated data reported by the RP of record by entering data items into the computer screener modules. The content for the two versions of the RP eligibility and the beneficiary capability screeners are described in the remainder of this section.

The computer selected the organization RP eligibility screener when the RP was an organization. It was designed to introduce the survey to the respondent, confirm the organization's contact information, determine the organization's and beneficiary's eligibility to be in the survey, and identify the most senior person who knew about the selected beneficiary's daily needs and how the selected beneficiary's Social Security funds were managed.

When the most senior person in the organization who knew about the beneficiary was identified, that person would finish the screener and would complete the RP interview. If the beneficiary had been selected to be interviewed, the beneficiary's contact information and legal guardian status were captured in the RP screener. If the organization was the beneficiary's legal guardian, the computer prompted the interviewer to obtain consent to contact the beneficiary; if someone other than the organization was the legal guardian, the computer prompted the interviewer to obtain contact information of the legal guardian.

The computer selected the individual RP screener when the RP was an individual person. As with the organization screener, the individual screener introduced the

survey, confirmed the RP's contact information, and determined the RP's and the beneficiary's eligibility to participate in the survey. Also, as with the organization screener, the individual RP screener determined whether the beneficiary had a legal guardian and if so, obtained contact information for the legal guardian so that consent to contact the beneficiary could be obtained.

Beneficiary Capability Screener

The purpose of the beneficiary capability screener was to introduce the survey, confirm the respondent's contact information, and determine whether the respondent was mentally capable of participating in the survey. The computer prompted the interviewer to answer two questions before launching the capability screener. The questions asked if the beneficiary lived in supervised group quarters and if the interviewer thought that the beneficiary might have problems completing the interview. The capability part of the screener was not always required; it was programmed to be skipped if not required.

The beneficiary capability screener was modified from a screener that was developed and validated for another Westat survey that had a similar population and was designed to determine the selected beneficiary's ability to understand and provide consent to participate in the survey. The beneficiary capability screener could be accessed in the computer only after the corresponding RP screener was complete and the beneficiary was eligible for the survey.

If the capability of the respondent needed to be determined, three questions were asked. The first question concerned recall about the purposes of the survey. The interviewer read the script that explained the purposes of the survey and then the interviewer asked the beneficiary to "recall" the purposes. If the beneficiary could not recall at least two of the three purposes of the survey, the interviewer explained the purposes again. The interviewer then asked the beneficiary a second time to recall the purpose of the survey. If the beneficiary could not recall at least two purposes of the survey the second time, the computer assigned the beneficiary a code of incapable and prompted the interviewer to begin identifying a proxy. The second and third questions concerned the beneficiary's understanding of the terms "voluntary" and "confidential" as used in the survey. As with the purpose of the survey, both of these two questions could be asked twice. If the beneficiary could not explain voluntary or confidential in two tries, the computer

coded the beneficiary as incapable and the computer prompted the interviewer to begin locating a proxy.

RP Interview Instruments

The RP interview instruments (for individual and organization RPs) were developed using draft topics provided by SSA. The interview instruments underwent extensive testing, as summarized in Section 3.1. Interview instruments used some "standard" questions (e.g., demographic characteristics) as well as questions that were developed and refined by Westat, NA, and the Committee. The RP interview instrument contained the following sections and content:

- Administration of payeeship – general experience with RP program and knowledge of SSA expectations of RPs;
- Administration of payeeship – specific to sampled beneficiary: relationship to beneficiary, frequency and type of contact with beneficiary, satisfaction and willingness to serve as the beneficiary's RP, geographical proximity to beneficiary, beneficiary's level of impairment (e.g., ability to communicate, reason for RP), and number of beneficiaries living with RP;
- Meeting needs – general satisfaction with understanding of beneficiary's needs and discussing beneficiary's needs with others;
- Meeting needs – specific to sampled beneficiary: the extent the beneficiary's needs have been met in the areas of housing, utilities, food, clothing, medical care, and finances; beneficiary's sources of income; and disagreement with beneficiary about how the SSA benefits were spent;
- Communicating about needs and payee involvement in meeting needs – needs included housing, food, clothing, medical care, and saving SSA benefits;
- SSA administration of representative payeeship – whether SSA informed RP of his/her responsibilities, RP recordkeeping and allocation of beneficiary's SSA funds, knowledge of SSA expectations of RPs, whether RP sought help from or reported information to SSA, RP use of SSA web site, and whether and why an RP–beneficiary relationship had ever been terminated;
- Demographics – age, education, race/ethnicity, income level and source(s), bankruptcy, residential transience, criminal history; and

- Alcohol and other drug problems – receipt or perceived need for treatment during past 5 years.

The RP interview instrument also aimed to identify types of RPs with the highest risk of misusing benefits. To this end, it measured the following topics: substance abuse history; transient residence; criminal history; prior RP status terminated; employment; and income and history of bankruptcy. Relevant to this aim was the RP's experience serving as an RP, such as whether the RP had ever had a beneficiary relationship terminated and why; whether the beneficiary's SSA funds were combined with another person's funds in an account; and whether the beneficiary had disagreed with the RP about how the beneficiary's SSA benefits were spent.

Beneficiary Interview Instrument

The beneficiary interview instrument also included some standard questions. Many of the topics and items in the beneficiary interview instrument were similar to those in the RP interview instrument, to facilitate comparison between the respondents. In addition, the beneficiary interview instrument included the following topics: living situation, interviewer rating of beneficiary, and direct indicators of exploitation/neglect. The beneficiary interview instrument contained the following sections and content.

- Living situation – type of residence, with whom, and duration;
- Administration of payeeship – number and duration of lifetime RPs, reason and need for RP;
- Administration of payeeship – specific to sampled RP: relationship to RP, frequency and type of communication with RP, satisfaction with RP, geographical proximity to RP, and knowledge of RP's responsibilities;
- Meeting needs – the extent the beneficiary's needs have been met in the areas of housing, utilities, food, clothing, and medical care, and disagreement with RP about how the SSA benefits were spent;
- Communicating about needs and payee involvement in meeting needs – needs included housing, food, clothing, medical care, and financial savings;
- SSA administration of payeeship – RP recordkeeping of beneficiary's SSA funds and SSA's level of helpfulness if beneficiary sought help;
- Demographic – age, education, employment, race/ethnicity;

- Alcohol and other drug problem – receipt or perceived need for treatment during past 5 years; and
- Validity check and interviewer rating – beneficiaries were asked whether and how RPs influenced responses, and interviewers recorded how well they thought the beneficiary understood the questions and was thinking clearly.

As the RP interview instrument aimed to identify types of RPs with highest risk of misusing SSA benefits, beneficiaries were also asked some related questions about their RPs. Specifically, beneficiaries were asked if basic needs were met and whether the RP influenced the beneficiary's interview responses.

3.1.4 Methods to Minimize Error

In questionnaire development, we used several approaches to minimize systematic and variable errors.

Systematic Errors (Bias). During questionnaire development and data preparation, we employed several measures to minimize systematic errors. To minimize under- and over-reporting, first, we electronically included range checks in the programming of the CAI. Second, we checked for and resolved most outliers in the data. Third, many of the response categories were truncated at the higher end (e.g., 12 months or more) or, if open-ended, they were collapsed to include infrequent responses in larger groups. Fourth, many item series (e.g., needs) were asked dichotomously; then, for positive responses, respondents were asked to estimate the frequency the event occurred – again, in broad categories.

Our efforts to minimize yea-saying (response acquiescence) involved the addition of "fictitious options" to some of the long strings of questions in both questionnaires. Data show a smaller yet still a high percentage of "yes" responses to fictitious options relative to "valid yes" responses. This response pattern suggests that some respondents responded to this interruption in valid responses to consider each response. However, the high endorsement rate of both types of yes responses suggest that some bias may have been involved in these response strings, which indicates a limitation of these questionnaires.

And finally, at the end of the beneficiary interview, the beneficiary was asked whether the RP told them how to answer any of the questions and if so was asked how. In addition, the interviewer rated whether he or she believed the interview was directly influenced by the immediate presence of the RP, how well the beneficiary understood the questions, and whether the beneficiary's responses were valid and reflected reality.

Variable Errors. To minimize order and context effects in the design of survey questions, the most sensitive items were put at the end of the questionnaire. These items include demographic information, history of bankruptcy and history of incarceration. We also ordered the questions from more general to specific.

3.2 Cognitive Testing of Interview Instruments

The RP and beneficiary interview instrument were drafted in late November 2005, then underwent extensive in-house expert review and review by the National Academies and the Committee. Following these reviews, the interviews underwent cognitive testing. The objectives of the cognitive tests were to observe the flow; learn how respondents understood the instruments; and to identify special situations that may need to be taken into account during RP and beneficiary interview administration.

3.2.1 Overview

From December 1 through December 8, 2005, Westat conducted 11 cognitive interviews to test the RP and beneficiary interview instruments.[4] Four beneficiaries participated in the test of the beneficiary interview instrument, and seven RPs participated in the test of the RP interview instrument. An incentive payment of 50 dollars was offered to respondents; in addition, respondents were reimbursed for travel costs and other expenses associated with completing an interview. Four cognitive interviewers from Westat conducted the interviews. All four were senior level interviewers with both formal training in cognitive interviewing and considerable hands-on experience.

Ten interviews were conducted at Westat's Focus Group/Usability Facility, and one interview was conducted off-site at the office of the participant. Standardized and ad hoc probes were added to the interviews to assist the cognitive interviewers. All interviews conducted at Westat were both audio- and video-recorded. The off-site interview was audio-recorded only. All interviews were observed by Westat, National Academies, and/or SSA staff. The on-site interviews were observed through a one-way mirror, and the off-site interview was attended by two observers in addition to the interviewer. During the interview, the interviewer recorded respondent responses. At the conclusion of the cognitive interviews, a debriefing was held. All interviewers and some observers attended and discussed responses and observations for both the RP and beneficiary interviews. Findings and recommendations for revising the instruments were summarized and delivered to the National Academies.

3.2.2 Respondent Recruitment

To conserve resources and complete the cognitive tests expeditiously, Social Security Administration (SSA) records were used to select 300 RPs and 300 beneficiaries from among those who lived within the vicinity of Westat's Rockville office. Both organization and individual RPs were selected. A letter sent to the selected RPs and beneficiaries explained the purpose of the cognitive testing and its importance to SSA, and encouraged recipients to participate. A toll-free number was given so potential participants could leave their names and numbers. The recruitment effort resulted in calls from 42

[4] The screening instruments did not undergo cognitive testing, however they were field tested.

potential participants to the toll-free number. Each person who expressed an interest in participation was called back and administered a screener. After the screening process was complete, the number of scheduled interviews was insufficient. To schedule additional interviews, the Westat recruiter telephoned additional names that SSA had provided.

In total, we conducted 11 cognitive interviews. The seven RP interviews were with one organization and six individual RPs (two parents, two other relatives, and two nonrelatives). This was consistent with our target for RP recruitment. Of note is that organization RPs were very difficult to locate. Only two were identified in the entire sample, and only one of those was interviewed.

Four beneficiary interviews were conducted. The target sample for beneficiaries consisted of at least two beneficiaries from each of the following categories: age 14 to 17 years, age 65 or older, a physical disability, and a psychiatric or cognitive disability. The actual beneficiaries interviewed consisted of two 15 year-old adolescents and two adults age 18 to 65. Both adult beneficiaries had a psychiatric disability and one also had a physical disability. No adults age 65 or over were found that were deemed competent to complete the interview. Beneficiaries older than 65 years of age were very difficult to contact. The sample contained 13 seniors. All 13 were contacted and all of them proved mentally incapable of participating in an interview. All 13 also lived in institutional settings, such as nursing homes and assisted living facilities.

3.2.3 Cognitive Test Findings

The RP and beneficiary interview instruments were modified to include cognitive testing probes to encourage respondents to indicate their understanding of the survey questions. As mentioned, cognitive interviews were completed with four beneficiaries and seven RPs. Respondents had diverse situations, which provided valuable input on how the questions would be understood by different types of respondents.

In general, most of the questions were understood by respondents. However, some of the sequencing of questions was confusing to respondents, so questions were reordered in the interviews to create pathways for different types of RP and beneficiaries. A few questions were awkward or did not appear to be applicable to some respondents, even though the questions were understood. Many of these questions were revised.

More substantial changes were made to the RP and beneficiary interview instruments in three instances. First, beneficiaries who were minors and living with parental RPs were confused by parental roles and RP roles, so some changes were made to skip patterns. Second, some RPs were paying an institution or facility where the beneficiary lived, and this made some of the benefit management questions difficult to answer. For example, some RPs could not break out housing and food costs because the institution provided both housing and food, and the costs will billed in one lump sum. Response categories were added for "don't know" and "not applicable" for these situations. When RPs responded "don't know" or "not applicable" to questions on how the benefit amounts were used, an open-ended question was added to probe for a reason why. Third, the RP and beneficiary relationship was different when the beneficiary and RP could not communicate. A series of questions on why the RP and beneficiary could not communicate was added. Depending on how this series was answered, a few questions about how often the RP and beneficiary communicated were skipped because the RP and beneficiary could not communicate. Once the revised RP and beneficiary interviews were drafted, additional materials and procedures were developed so that a field test could be conducted.

3.3 Field Test Activities

A field test was conducted in January and February of 2006. The purpose of the field test was to test the draft field procedures, field materials, and survey instruments. At the time of the field test, the Office of Management and Budget (OMB) clearance had not been received, so data was collected from no more than nine respondents for each instrument.

Social Security Administration (SSA) records were used to select 50 beneficiaries and their RPs from among those who lived within the vicinity of Westat's Rockville office. Beneficiaries were selected based on several factors, including living arrangements, age, and whether the RP was an individual or organization.

Advance letters were mailed to 84 individual or organization RPs prior to the start of field contacts. These letters, which were on SSA letterhead and signed by an SSA official, introduced the Survey, described its purpose, requested participation in the Survey, and indicated that an interviewer from Westat would be contacting the recipient shortly.

The field test was conducted in two phases. The first phase was conducted between January 5 and January 16, 2006, and focused on individual RPs and their beneficiaries. The second phase was conducted between February 2 and February 13, 2006, and focused on organization RPs and their beneficiaries.

Interviewers were trained for each phase of the field test. After completing a training session, the interviewers were assigned to paired RP and beneficiary (dyad) cases. The interviewers contacted, screened, and interviewed respondents and recorded findings. Instruments used in the field test were administered using paper and pencil. During the first phase, interviews were completed with nine individual RPs and seven beneficiaries. In the second phase, the interviewers contacted seven organization RPs and completed three interviews. Beneficiaries in two of the organizations were contacted and screened, but neither passed the capability screening.

Based on the results of the field test, discussions with the National Academies and comments from SSA, Westat revised field procedures, materials, and survey instruments for use in the main data collection effort.

3.4 Translation

The RP and beneficiary screeners and interviews were translated into universal Spanish by professional translators. The translation was reviewed by up to five editors as well as by bilingual Westat staff who were heavily involved in the development of the screeners and interviews. Bilingual field interviewers were trained to administer both the English and Spanish versions of the survey instruments. A few minor irregularities in the translation were discovered after the Spanish instruments were programmed (e.g., a probe that was not translated into Spanish and typos, such as an extra or incorrect word and incorrect verb tenses). These irregularities were documented and explained to the bilingual interviewers at training.

3.5 Development of Computer-Assisted Interviewing Applications and Associated Systems

Blaise for Windows software, which operates on laptop computers, was used on the survey to collect the RP and beneficiary data. The Blaise system has been developed by Statistics Netherlands and is widely recognized internationally and in the United States as the leading CAI system for performing research surveys like those most often sponsored by government agencies.

Senior members of the Blaise programming team worked with instrument designers during the screener and questionnaire design process. The programming team provided advice and answered questions about the software's capabilities and the relative resource requirements of alternate design strategies. This coordination facilitated the design and development processes, and helped ensure that high-quality instruments were programmed and tested within the tight project schedule.

Westat's SpecWriter software was used to enter and maintain the programming specifications. SpecWriter produces a hard-copy version of the specifications and also generates Blaise block templates. Since a different system had been used to produce specifications and edits for the survey's field test (which was conducted using paper questionnaires rather than CAI), a customized program was developed to import these specifications into the SpecWriter database, thus expediting the process for sections and items that did not change between the field test and the main data collection effort.

Once each section was specified, the SpecWriter output was given to a Blaise programmer to write the Blaise code. The coding process was greatly facilitated by the Blaise templates that SpecWriter produced, which contained the text for all of the question and answer categories.

Each questionnaire was tested at both the section level (unit testing) and overall (integration testing). Testing specialists used the detailed specifications produced by SpecWriter to compare data entry screens and functions against requirements and expected results. All problems were entered into a computerized tracking system so that progress in fixing both design errors and programming bugs could be monitored. Questionnaire design staff carried out final acceptance testing.

Since Blaise is an integrated system for data editing as well as data collection, staff used it for both purposes on the survey. Many edits were included in the data collection programs on the laptops, so that questionable or inconsistent data could be flagged and reviewed immediately with the respondent, rather than being first discovered long afterward during home office data editing. These included "hard" range checks ("impossible" values that must be corrected with the respondent before the interview can proceed) and "soft" range checks (unlikely values that may be either accepted or corrected after double-checking with the respondent).

Although building such edits directly into the interviewing process is desirable, an excessive number of edits can be confusing to respondents (particularly those with cognitive impairments) and interviewers, and may disrupt the overall flow of the interview. Thus, content experts and staff with expertise in interviewing difficult populations evaluated the proposed edits in order to properly balance all of the relevant considerations.

The survey data were maintained in Blaise throughout the data editing process. When changes to the data were necessary (based on interviewer remarks, for instance), the editing staff made and documented the changes. Once changes were made to a particular case, the skip pattern and range edits were rerun to make sure that the data were still clean.

4. Data Collection

This chapter contains a discussion about the recruitment and training of field staff, as well as interviewing activities, data security, data editing and cleaning activities, quality control efforts during data collection, and survey response rates.

4.1 Recruiting Activities

The survey design required that 9 field supervisors and approximately 124 interviewers be hired. Supervisors were hired in January 2006 and interviewer candidates were hired between April and early July 2006. The recruitment process ended about 3 weeks before training, to allow time to receive security clearance for the field staff.

4.1.1 Field Staff Recruitment

Nine field supervisors were chosen to supervise interviewers within nine field areas. The supervisors chosen from Westat's supervisor pool were all experienced Westat field supervisors who were highly qualified and motivated and who were available to work full time during the 16-week data collection period. Some of these field supervisors were brought on the study prior to the data collection period to assist home office staff with the recruitment of interviewers.

The survey design was based on hiring between one and three interviewers for each of the 60 Primary Sampling Areas (PSUs), depending on the sample size within the PSU. The goal was to hire about 124 interviewers who would work at least 20 hours a week over the data collection period. The goal included about 20 extra candidates to accommodate interviewer attrition (i.e., when interviewers terminated earlier than expected for any reason).

Initially, interviewers were recruited from Westat's pool of experienced data collectors. Additional candidates were recruited from recommendations of experienced Westat interviewers and from classified newspaper advertisements and specific internet sites. These candidates were screened for communication skills, computer skills, and long- and short-time travel availability. Some candidates would be used to travel to "problem" PSUs to serve as troubleshooters and refusal conversion specialists. In addition, Spanish language candidates were screened by bilingual project staff for their ability to communicate effectively in both Spanish and English. Approximately 9 percent of the total candidates hired were bilingual. Approximately 70 percent of the English and bilingual candidates had prior experience relevant to data collection.

4.1.2 Field Staff Attrition

In total, 134 interviewer candidates were recruited for the Survey of Social Security Representative Payees and Beneficiaries (Survey). Since recruitment was conducted over a 3-month period, some candidates dropped out and some were replaced prior to the end of recruiting. Of these candidates, 111 were expected at training. Some attrition occurred just before and during training; and 105 (95%) candidates successfully

completed training. During the field period, eight additional interviewers resigned or were terminated prior to completing their assignment. Total interviewer attrition for the Survey was 13 percent.[5]

4.2 Training Activities

Three formal training sessions were held prior to data collection: train-the-trainer, supervisor training, and interviewer training. Informal training continued throughout the data collection period as needed.

4.2.1 Train the Trainer and Supervisor Training

About 1 week prior to interviewer training, a train-the-trainer session and a supervisor training session were held at Westat. Nine field supervisors and 19 home office trainers attended a 3-day in-person train-the-trainer session. The purpose of this session was to familiarize the training staff with survey materials and their specific roles at training. Trainers were supplied with survey-specific procedures, manuals, and trainer guides. The guides provided a compendium of all materials required for training, including detailed training agendas, procedures and materials to be presented in each session, scripts for all lectures and practice role plays, exercises, and answer keys for all role plays and exercises.

After trainer training, all field supervisors attended a 2-day training session to learn about their roles and responsibilities.

4.2.2 Interviewer Training

All trainees participated in a 5½-day interviewer training session. Trainees new to Westat attended an additional half-day training on general interview techniques.

[5] Total attrition is defined as the percentage of interviewers lost between the number expected at training (111) and the number completing their assignments (97).

Bilingual trainees also attended an additional half-day training that concentrated on reviewing bilingual scripts and materials.

During training, concurrent training sessions for smaller groups (which consisted of 2 trainers, 2 training assistants, and approximately 16 trainees) were held in multiple rooms. Lectures and role play practice sessions were scripted to guarantee that all trainees were trained on the same material. During training, all trainees were trained to administer the survey instruments in the same prescribed way. Trainees were supplied with a project-specific procedures manual, a computer user guide, and a question-by-question specification guide for the survey instruments. Throughout the training, trainees were placed in situations where they had to use the procedures and survey instruments as they would when actually collecting data.

Trainers met at the end of each training day to evaluate trainees on their level of confidence and competence. If trainees were thought to need help, they were asked to attend remedial or practice labs that were scheduled most evenings. All trainees could sign up for these labs. In labs, trainers reviewed sections of the manual, supervised practice role plays and exercises, and provided help with computer matters. On the last day of training, trainees participated in a certification session that required them to role play the entire survey process. Trainers determined that a few trainees needed some additional practice prior to starting data collection, and these trainees worked with their field supervisors until they were considered competent.

4.3 Implementation of Data Collection

Data collection was completed in a 16-week period between late July 2006 and mid-November 2006. Attempts were made to contact and interview approximately 500 organization representative payees (RPs), 4,600 individual RPs, and about 2,550 beneficiaries. RPs were administered an RP screener to determine the eligibility of both the RP and the beneficiary. Under certain circumstances, beneficiaries were screened to determine whether they were capable of participating in the survey.6 If they were not capable, the interviewer would try to locate a proxy. To be an acceptable proxy, the person

[6] Beneficiaries were screened for capability to participate if the RP volunteered that the beneficiary might not be able to complete the interview, the beneficiary was an adult with a legal guardian, the beneficiary lived in supervised group housing, or the interviewer had reason to believe the beneficiary might not be able to complete the interview.

had to be an adult (18 years or older) and could not be the RP or the spouse of the RP. Interviewing was completed using computer-assisted interviewing (CAI) on laptop computers. Survey instruments were available in English and Spanish. Interviewers conducted 130 interviews in Spanish, of which 82 were for RPs and 48 were for beneficiaries or proxies. If RP respondents did not speak English or Spanish, the interviewer attempted to locate a translator; beneficiary and proxy interviews were not completed using a translator.

The remainder of the section discusses the steps that were taken to locate, contact, and interview respondents.

4.3.1 Sending Advance Materials

Approximately 2 weeks prior to contact, advance letters were mailed to beneficiaries, organization and individual RPs, and parents or legal guardians of minor beneficiaries. Advance letters signed by a Social Security Administration (SSA) official were reproduced on SSA stationary and mailed in SSA franked envelopes. The letter provided an introduction to the survey, information on confidentiality, and notice that an interviewer would be contacting the RP or beneficiary. The RP letter also indicated that the RP must complete the interview; the beneficiary letter indicated that beneficiary participation was voluntary. Each letter included a Privacy Act statement and a Paperwork Reduction Act statement. The letters are in Appendix B.

Each envelope was stamped with a change of address request, and a Westat post office box was set up to receive returned mail and change of address information.

4.3.2 Contacting the Respondent

For several reasons, interviewers were instructed to contact RPs first before attempting contact with a beneficiary. First, since the RP interview was mandatory, the RP was more likely to complete the interview. Second, the RP determined the eligibility of both the RP and the beneficiary. Third, the RP could give the interviewer important information about the beneficiary, including whether the beneficiary had a legal guardian, and information on how and when to contact the beneficiary. Finally, the RP might volunteer

information about the beneficiary's capability of participating in the Survey and (if needed) assist in identifying a proxy.

Interviewers were trained to contact and conduct screenings and interviews with RPs by telephone whenever possible. When contact was not possible by telephone, RPs were contacted in person. If the RP was ineligible for the survey, the beneficiary was not approached.

Interviewers were allowed to contact beneficiaries to set up appointments for interview, either in person or by telephone. If initial contacts were made in person, interviewers would hand beneficiaries brochures to help establish rapport and the legitimacy of the Survey. Capability screening and interviewing of beneficiaries was always conducted in person using CAI. The brief capability screening was administered to determine whether the beneficiary was incapable of participating in the survey for himself or herself, for example, due to severe mental or developmental impairments. If the beneficiary failed the capability screening, the interviewer attempted to locate a proxy.

When telephone numbers were available, interviewers were required to make up to five telephone attempts to contact an RP within a 7-day period, before attempting to make contact in person. When telephone contact was not possible or successful, interviewers were required to make up to up to five in-person visits to the address within a 2-week period. When no one was at the address on the first visit, interviewers were instructed to visit neighbors to determine the best day and time to contact the RP. When RPs were no longer living at the address, the interviewer began field tracing. (These requirements also applied to beneficiaries, depending on the beneficiary's situation.)

4.3.3 Determining Eligibility

Interviewers administered screenings to RPs named in SSA administrative records to determine the eligibility of both the RP and the beneficiary. To be eligible for the survey, an individual RP had to represent no more than 14 beneficiaries at the time of screening; had to be serving as the RP for the reference beneficiary[7] on Sample Reference

[7] During sample selection, a reference beneficiary was selected for each RP. The RP answered questions about the reference beneficiary during the interview. One-half of the reference beneficiaries were also selected to complete an interview.

Date (SRD) (i.e., January 1, 2006), or at the time of screening; and the reference beneficiary had to be receiving payments from SSA Title II or Title XVI on the SRD. For an organization RP to be eligible for the survey, the organization could not receive a fee for service for serving as the RP; had to represent no more than 49 beneficiaries at the time of screening; had to be serving as the RP for the reference beneficiary on SRD or at the time of screening; and the reference beneficiary had to be receiving payments from SSA Title II or Title XVI on the SRD. As with RPs, beneficiaries were eligible to participate in the survey if they met all of the criteria mentioned above and were at least 14 years old on SDR.

4.3.4 Obtaining Consent/Child Assent

Prior to conducting the interview, interviewers asked RPs, adult beneficiaries without legal guardians, and proxies to give verbal consent to participate in the survey. In the 121 completed cases where the adult beneficiary had a legal guardian or the 294 completed cases where the beneficiary was a minor, the interviewer had to obtain consent from the parent or legal guardian for the beneficiary to participate in the Survey, prior to contacting the beneficiary. The consent could be obtained either verbally over the telephone or in writing in person. For the 294 cases where the beneficiary was a minor, interviewers always asked for written assent to participate in the survey from the minor, in person, prior to administering the interview. The consent/assent forms are in Appendix B.

4.3.5 Locating Respondents

The first step in the data collection process was to determine whether the location information supplied by the Social Security Administration (SSA) for sampled persons was up to date and accurate. If not, prefield tracing activities at the home office were undertaken. When interviewers could not locate sampled persons from location information provided from SSA administrative records or from prefield tracing activities, interviewers undertook field tracing activities. If field tracing efforts did not locate the sampled person, post-field tracing was undertaken.

Prefield Tracing Activities

Once the sample was selected, prefield tracing activities were started by searching two electronic databases for more current information. First, RP and beneficiary address files were sent to a vendor that searched for any changes of address, using the National Change of Address (NCOA) database provided by the United States Postal System. All addresses were then sent to another vendor that linked the most recent telephone number with the addresses. When new addresses and telephone numbers were found, that information was sent to Westat's Telephone Research Center (TRC) for verification. The TRC conducted prefield tracing for all organization RPs and about 20 percent of the individual RPs.

TRC tracers contacted individual RPs to verify their name and contact information. Because the location information supplied by SSA for organization RPs included only the name of the organization and not a contact name, all organization RPs were contacted. During the call, TRC tracers confirmed the name and contact information of the organization, and attempted to identify the name and title of the appropriate person in the organization who could respond to the RP screening and interview.

If the TRC tracer could not locate the RP using the new location information, the tracer called the original telephone number. If the original number was not productive, the tracer called the beneficiary to try to locate the RP. Other TRC tracing activities included calling directory assistance, reverse directory assistance search, and use of a variety of internet search engines. The results of cases sent to prefield tracing are shown in Table 4-1.

Table 4-1. Prefield tracing results

Result	Number of Cases	Percent of Cases
Found	1,201	74
Not Located	240	15
Other Nonresponse	173	11
Total	1,614	100

Field Tracing Activities

When location information supplied to the interviewer was not helpful in locating the RP, interviewers conducted field tracing activities. These activities included attempting to contact the beneficiary by telephone to locate the RP, and contacting directory assistance for new RP location information. Next, the interviewers conducted in-person visits to attempt to locate the RP. During these visits, interviewers contacted the current resident at the RP address and contacted neighbors near the RP address to obtain location information. Up to three in-field contacts of the current resident or neighbors were made to locate the RP. When a person with the same name as the RP was found, the interviewer confirmed the identity of that person by using known birth date and gender information, and by confirming that the person managed Title II or Title XVI payments for one or more individuals.

Post Field Tracing

If field tracing activities failed, TRC home office staff or field supervisor staff queried internet search engines to obtain leads. Search engines queried included theultimates.com, zabasearch.com, google.com, and verizon.com. If leads were found, home office or field supervisor staff followed up with leads to locate the RPs and then contacted RPs to verify their identity and confirm their location information prior to sending the information back to an interviewer. The results of cases sent to post field tracing are shown in Table 4-2.

Table 4-2. Post field tracing results

Result	Number of Cases	Percent of Cases
Found	12	10
Not Located	108	90
Other Nonresponse	0	0
Total	120	100

4.4 Data Security

The security of all data associated with the Survey was a major concern before, during, and after the data collection phase. This section summarizes some of the major security concerns and precautions. Greater detail can be found in the project's Data Security Plan, which was submitted to the National Academies in May 2006.

Because data security is a significant issue in nearly all Westat projects, the company has developed a corporate Information Technology and Systems (ITS) Security Policy and Best Practices document, which is the basis for the security procedures and practices on all Westat studies. This section discusses corporate policies and capabilities in the areas of facility and computer security, data security, network and data communications security, personnel security, disaster recovery, and user assistance and incidence reporting. Corporate practices include the following.

- Controlled physical access to all Westat facilities and computer centers.
- Password-protected accounts and role-based restricted access to all computer systems and data.
- Strong encryption to protect sensitive data during transmission and when stored on devices (such as laptop computers) located outside the computer centers.
- A system of firewall-protected network security zones to ensure the maximum possible protection for all data.
- A requirement that all Westat employees sign an assurance of confidentiality and receive instruction in Westat's data security policies, standards, and procedures.
- Systems to protect computer facilities from fire, power surges, electrical outages, and excess heat, along with a formal disaster recovery plan to be used in the event of a significant failure of regular computing services.

Two areas of particular concern to the Survey were the protection of data received from the Social Security Administration and the protection of data stored on laptop computers. These concerns are discussed in the next section.

4.4.1 Protection of Social Security Administration Data

Westat received administrative data files from SSA to select the sample of RPs and beneficiaries who would be interviewed for the Survey. Because these data files contained individually identifiable information, special precautions were taken to protect the confidentiality of the data. The CDs and DVDs that were received from SSA were kept in locked file cabinets at all times when not in use, and were destroyed prior to the end of the contract. Access to these data were restricted to a subset of project staff who needed such access to perform their assignments. Social Security numbers were not included in the project's field management system database and were never stored on laptop computers.

In addition to these precautions, the data were protected by all standard Westat security systems and practices. Access to personal computers and network data storage areas were password protected, and no data were stored on local PC hard drives or removed from Westat offices. The network security zone used to execute the sampling programs was protected by a programmable firewall and was not accessible from the public Internet. All staff working on the project were required to read and pledge compliance with Westat's "Employee or Contractor's Assurance of Confidentiality of Survey Data." In addition, all project staff were required to undergo National Agency Criminal Investigation (NACI) security clearance.

4.4.2 Protection of Data Stored on Laptop Computers

Interviewers collected data from RPs and beneficiaries using CAI. All data were collected on laptop computers using Blaise for Windows software running under the Windows XP Professional operating system.

Management of the data collection activities was provided by the Field Operations System (FOS), a customizable Westat corporate system that provides a standardized management system framework within a secure computing environment for field data collection projects. FOS functions included case assignment, case status management, and data transmission. FOS applications were written using Visual Basic, Active Server Pages (ASP), and Crystal Reports; commercial off-the-shelf software (COTS) was also used to provide certain functionality, such as secure data transmission. Microsoft SQL Server was used to store FOS data both on the laptop and at the home office.

Access to the laptops was protected by a power-on username and password that were unique to each interviewer. Interview data were stored in files that were encrypted using a third-party software package (Entrust). In addition, the directories that contained interview or FOS data were encrypted using Windows Encrypting File System (EFS).

A private dial-in network was used to move interview and management data between the laptops and the home office. The data files remained in encrypted form during the data transmission process, which was also protected through the use of Secure Socket Layer (SSL). The home office servers where the transmitted data files were stored was installed on a special subnetwork that was isolated from other portions of the Westat network by a firewall, so that access to these systems and data was limited to those users who were directly involved in field operations support.

4.5 Data Editing and Cleaning

Data editing and cleaning encompassed coding and editing procedures within survey instruments and trace outside of survey instruments.

4.5.1 Coding and Editing Procedures within Survey Instruments

Data collection for the Survey was conducted using Blaise software, which was used to develop CAI instruments. Blaise is an integrated system used for data collection as well as data editing. Blaise provided both "interviewing" and "data editing" access to the survey data. Acceptable ranges and consistency checks were programmed into the RP eligibility screener, beneficiary capability screener, RP interview, and beneficiary/proxy interview. The interviewing mode of Blaise allowed full navigation control to the interviewer, but required that hard consistency checks be resolved before moving forward, and also required that the route (order of questions) be maintained.

Range Specifications

Across all types of instruments, responses were subjected to both "hard" and "soft" range edits during the interviewing process. A "hard" range check is a logical inconsistency so inconceivable that it must be reviewed and corrected with the respondent immediately. With a hard edit, the interviewer could not continue to the next field until the problem had been resolved. Question 27a in the RP interview is a simple example of this: "How much pocket money did [BENEFILL] get from his or her Social Security benefits?" The allowable dollar amount for this variable is $1 to $2,050; a response less than $1 or more than $2,050 would not be permissible. Whenever a hard range was triggered, the interviewer saw a screen describing the items that were inconsistent, and selected one item to re-ask and re-enter. By definition, the interview could not proceed until the hard range check had been satisfied. If the respondent or interviewer insisted that a response outside the "hard" range was valid, the interviewer was instructed to enter the information in a comments field.

With a soft edit, the interviewer could either change the value of the field or suppress the edit. Soft ranges are designed to point out unusual or unlikely situations so that the interviewer may verify them with the respondent. The same items are shown on the screen, but the interviewer has an additional option to proceed with the interview without changing any data items. An edit may be made "soft" in data collection mode and "hard" in data editing mode.

Consistency Checks

Consistency checks or logic edits examined the relationship between responses to ensure that they did not conflict with another response, or that the response to one item did not make the response to another item unlikely. Questions 9z and 9 in the beneficiary questionnaire are examples of simple consistency checks. Q9z asks: "Are you related to [REPFILL]?" If the beneficiary responds, "No." a consistency check enforced that Q9. "What is [REPFILL'S] relationship to you?" was not asked and prevented contradictory or illogical relationship responses. If the verified response still resulted in a logic error, or the response to a primary question was incorrectly entered (and the interviewer could not back up to the respective question), the interviewer recorded the problem either in a comment in the CAI instrument or on a problem report.

4.5.2 Coding and Editing Procedures Outside of Survey Instruments

Not all data errors could be caught and corrected within the coding and editing specifications incorporated in the CAI instruments. Data management staff reviewed comment fields, other specifies, problem/communication sheets, frequencies, cross-tabulations, and data reports to discover and correct discrepancies and anomalies.

Editing from Instrument Content

Additional reports were developed to check data validity and ensure that the Survey protocol was followed. Also, hard copy communication/problem sheets that were sent to the home office in case folders were reviewed.

In the case of the RP eligibility screener, a report was developed that compared the name of the sampled RP to the name of the RP that was entered in the screener. If the report indicated that the interview was not conducted with the sampled person of record, the data were excluded from the analysis files, a nonresponse disposition code was assigned, and the case was documented in the unit decision log.

If a proxy completed the beneficiary questionnaire, another report compared the date of birth of the beneficiary from that interview to the beneficiary's date of birth on the SSA files. Discrepant birthdates were reviewed and, if two out of three components of the birth date did not match (month, date, or year), analysts determined that the proxies most likely became confused in the demographic section of the interview and incorrectly responded to these questions about themselves instead of the beneficiary. In these instances, analysts decided to remove the demographic data from the analysis files. Such cases were also documented in the unit decision log.

Communication/Problem sheets submitted in case folders were also reviewed for data updates. While most of this type of hard copy did not result in updates, some did. These instances usually occurred when the cases were closed and either the interviewer forgot to enter a comment or the respondent offered additional information that may have affected a response in the interview. Thus, all communication/problem sheets were

reviewed either for updating or to notify the field staff of special circumstances. In situations where data management staff determined that an update was needed, the change was entered in a transaction file that recorded the variable name, the original value, the updated value, the user, and the date and time.

Frequency and Cross-Tabulation Review

SAS programs were developed to perform edit checks on the interview data. All interview skip patterns were checked to ensure that data did not exist for data items that should have been skipped and that data values were missing only when a data item had been properly skipped. After the SAS edits were reviewed and the appropriate updates applied, frequencies were produced for all variables. Each variable was reviewed and evaluated to ensure the accuracy of the Blaise programs and integrity of edited data. Experienced data specialists reviewed frequencies to identify outliers, unexpected missing data, and data inconsistencies. When a potential problem was identified, the data manager located the corresponding variables, identified the case, and evaluated the data to determine if any updates were needed. Frequencies were run and posted weekly to the project's quality assurance site. (See section 6.2 for a description.)

Review of Other Specify and Interviewer Remarks

Data management staff reviewed remarks entered electronically by interviewers and questions that had "other specify" as a response choice (i.e., questions that had too many multiple response options to list). The review of interviewer remarks took into consideration the content of the comment and at what point in the interview the remark was inserted. If the remark stated a response was entered in error or the remark offered an explanation that changed the response, data management staff entered an update in a transaction file that recorded the variable name, the original value, the updated value, the user, and the date and time.

Data management staff reviewed questions such as "How is the Proxy related to [RepPayee]?" that offered "other" as a response choice and then asked the respondent to explain what was meant by *other*, to see if any of these responses would fit into existing response categories associated with that particular question. If the response could be coded

into an existing category, the change was entered into the transaction file, requiring variable name, old value, new value, user, and date and time.

Review of Help Desk and Audit Trails

Data management staff reviewed reports from the help desk and audit trails to identify data issues or discrepancies, and then applied updates in the home office where applicable. Data items that could be corrected in the home office from help desk reports followed along the lines of the communication sheets and interviewer remarks (i.e., new or different information that was obtained or relayed after the interviewer closed out the case). Updates were recorded in the transaction files. The audit trail (a keystroke-by-keystroke record of all responses entered during the CAI, and the date and time that each response was entered) was used when data were incorrectly stored in the database or an unexpected interruption occurred during the CAI. Audit trails were used for reconstructing complete or partial interviews keystroke by keystroke, or for just retrieving one or two responses.

4.6 Quality Control Efforts During Data Collection

This section contains information on the quality control efforts associated with the data collection effort, namely quality control during and after training and validation of the data collected. Other quality control efforts are discussed in other sections of this report; for example, instrument and field testing are discussed in Chapter 3, data processing quality control is discussed earlier in this chapter, and quality assurance reporting is discussed in Chapter 6.

4.6.1 Quality Control During and After Training

Personnel chosen to act as trainers for interviewer training were highly qualified project staff and field supervisors, almost all of whom had been part of training teams for large interviewer training sessions in the past. As mentioned, training personnel underwent a structured 3-day in-person train-the-trainer session at Westat to learn survey-specific procedures, scripted training materials, and how to use computer equipment such

as data displays. The in-person 5½-day interviewer training program helped ensure consistency in data collection by using scripted lectures and standardized exercises, with a heavy focus on scripted practice sessions. During training, trainees were given many practice role plays using a training database. For each screening and interview, a demonstration session was performed by trainers, followed by a scripted interactive session where the trainer played the respondent and trainees took turns being the interviewer. This interactive session was followed by two role play sessions in which two paired trainees took turns being an interviewer or a respondent. All role play sessions were scripted so that the person playing the respondent answered the questions in predetermined ways. Prior to the conclusion of training, trainees were paired and each completed a certification script that entailed each trainee contacting the mock respondent, conducting the screening and interview, ending the contact, and completing all the paperwork for the case. Trainers observed this process and assessed performance on standardized evaluation forms.

Trainees were given assignments of additional practice sessions with supervisors and family members after training, if trainers thought that they needed more practice prior to starting data collection. Every 2 weeks during the data collection period, the home office supervisory staff and the nine field supervisors had conference calls to discuss any problems the supervisors were having and any revisions to procedures the home office staff had implemented. These conference calls helped ensure consistency in supervisory actions and discussions with interviewers in each of the nine areas. Home office staff followed up decisions and revisions with supervisors in periodic email messages during the data collection period, and interviewers were sent periodic field memos discussing identified problems and suggested solutions, as well as reminders of key protocol points to follow.

4.6.2 Validation of Interviews

Validation selection of completed interviews was based on three goals: (1) select an adequate number of RP and beneficiary interviews for each interviewer to maintain data integrity; (2) select an interview from each interviewer's first few completed interviews; and (3) validate interviews completed throughout the duration of the data collection period.

Selection of Interviews

Three methods were used to select interviews for validation: (1) selection based on predetermined date, (2) selection based on length of interview, and (3) other purposive selection. The criteria used for selection, by predetermined date, included selecting the third interview the interviewer completed regardless of respondent type (i.e., RP or beneficiary). If fewer than three interviews had been completed by the end of the first week of interviewing, the last interview completed before that date was selected. If no interviews had been completed by the end of the first week of interviewing, the first interview completed thereafter was selected for validation. The remaining four interviews were selected based on specific predetermined dates in August, September, October, and November and required a specific type of respondent.

Criteria used for selection based on length of interview were RP interviews taking 10 minutes or less, beneficiary interviews taking 5 minutes or less, and proxy interviews taking 6 minutes or less. All interviews falling into these criteria were sent to the Telephone Research Center (TRC) or to field supervisors for validation.

Other purposive selection resulted when home office supervisory staff or field supervisors deemed some aspect of the interview or interviewing process to be questionable (e.g., results of an earlier validation were questionable); or when supervisory staff had some other problem with an interviewer's work.

Telephone and Mail Procedures

Initial attempts to validate interviews were first made by telephone. Validations were completed by TRC validation staff or by field supervisory staff. An electronic file was sent to the validator with information from the interview that would be used in the validation. The validator telephoned the respondent and administered a brief validation questionnaire. Questions included whether the respondent remembered the interview (if not, the interviewer confirmed the telephone number and name of the respondent and told the respondent what the survey was about and the interview date to help enhance recall); whether the survey was conducted on the telephone or in person; whether the interview was conducted on a laptop computer; whether the interviewer was polite; who the respondent answered questions about and how the respondent was related to the RP or

beneficiary; highest level of education RP or beneficiary attained; how long the interview took; and whether the respondent would like to comment further about the interviewer or the survey. Slight wording changes were used for these questions depending on the type of respondent (RP, beneficiary, or proxy).

When respondents selected for validation could not be reached by telephone, a letter was mailed to the respondent with a brief questionnaire and a self-addressed, postage paid return envelope. Based on information found during the validation contact, an outcome was determined and classified as acceptable, questionable, or unacceptable.

Validation Results

Outcomes were recorded in a validation system that was used for tracking the status of cases and the results by interviewer. If the validator determined that the interviewer was not implementing procedures correctly, the interviewer was given corrective feedback by supervisory staff. If the interviewer was suspected of falsification, all of the interviewer's completed interviews were selected for validation and the interviewer was told to stop work while an investigation was completed. When supervisory staff determined that the interviewer had falsified information, the interviewer was terminated; the data from the falsified interviews were removed for the data sets (except for two interviews that were identified after the datasets were finalized, which are documented in the unit decision log); and if time permitted, the interview was reassigned to another interviewer to be completed.

In total, 822 (13.5%) of the completed interviews were selected for validation. (See Table 4-3.) This includes 293 (16.5%) of the interviews for beneficiaries/proxies and 529 (12.3%) of those for RPs. Of those interviews selected for validation, 15 (5.1%) of the beneficiary/proxy interviews were judged to be unacceptable, and 12 (2.3%) of the RP interviews were unacceptable. Three interviewers accounted for the unacceptable interviews identified through validation.

Receipt Control

Cases folders returned from the field were receipted in the receipt control system on a flow basis. Data management staff recorded the final disposition code for each type of instrument (i.e., RP eligibility screener, RP interview, beneficiary capability screener, beneficiary interview) and that the case information sheet was returned. Hard-copy consent and assent forms were also receipted, if applicable. Data management staff reviewed the contents of each folder for communication sheets, unusual event forms, and other correspondence from the field. As needed, cases were flagged for home office supervisory staff review, to resolve inconsistencies in field procedures.

Table 4-3. Interview validation results by type of respondent

	Type of Respondent		
	Beneficiary	RP	Total
Interviews Completed	1,771	4,297	6,068
Interviews Selected for Validation	293	529	822
% Selected for Validation	16.5%	12.3%	13.5%
Validation Acceptable	239	454	693
Validation Unacceptable	15	12	27
Non Response	39	63	102
% Acceptable	81.6%	85.8%	84.3%
% Unacceptable	5.1%	2.3%	3.3%
% Nonresponse	13.3%	11.9%	12.4%

The receipt control system reflected the final disposition codes for each survey instrument. This system was used to reconcile the final disposition codes with the field management system; it served as the basis for calculating response rates for the survey.

4.8 Response Rates

Response rates measure aspects of both survey procedures and respondent behavior. Unweighted response rates are often used as survey process measures, which are of interest to survey managers; whereas, weighted response rates are often used as indirect measures of data quality, which are of interest to data users. Table 4-4 contains weighted and unweighted eligibility rates and response rates for the RP eligibility screener and RP interview. Table 4-5 contains the rates for the beneficiary interview.

The eligibility rates in Tables 4-4 and 4-5 were calculated as follows:

$$eligibility\ rate = ER = \frac{C_{ER} + C_{EN}}{C_{ER} + C_{EN} + C_{FI}},$$

where

C_{ER} = count of *eligible respondents* (i.e., RPs or beneficiaries who the interviewer had determined were eligible and who had also completed the interview questionnaire);

C_{EN} = count of *eligible nonrespondents* (i.e., RPs or beneficiaries who the interviewer had determined were eligible but who had *not* completed the interview questionnaire); and

C_{FI} = the sum of the base weights for *field-determined ineligibles* (i.e., RPs or beneficiaries who the interviewer had determined were ineligible).

The response rates in Tables 4-4 and 4-5 were calculated as follows, using response rate formula #4 recommended by the American Association for Public Opinion Research (2006):

$$response\ rate = RR4 = \frac{C_{ER}}{C_{ER} + C_{EN} + (ER)(C_U)},$$

where

C_U = count of *nonrespondents of unknown ineligibility* (i.e., RPs or beneficiaries who had completed neither the eligibility screener nor the interview questionnaire and hence their eligibility status was unknown).

For the *unweighted* eligibility and response rates, the quantities C_{ER}, C_{EN}, C_{FI} and C_U—referred to as *status counts*—are unweighted counts; whereas for the weighted rates, they are the sums of base weights. (Base weights are discussed in Chapter 5.)

The weighted and unweighted status counts were calculated by aggregating weighted and unweighted counts by final disposition code. Table 4-6 contains unweighted, entire-sample-level counts by final disposition code for the RP eligibility screener, and it indicates how disposition-code counts were aggregated to status counts. Tables 4-7 and 4-8 contain corresponding information for the RP interview and the beneficiary interview,

respectively. Although response rates were not calculated for the beneficiary capability screener, Table 4-9 lists the final disposition codes and associated unweighted counts for the beneficiary capability screener.

Some of the response-rate patterns observable in Tables 4-4 and 4-5 are the following.

- Within each of the various domains, the response rate for the RP eligibility screener and the RP interview were approximately the same. The unweighted response rate for all RP eligibility screeners was 92 percent and that for all RP interviews was 91 percent.
- Within each of the various domains, the response rate for the beneficiary interview was lower than that for the RP screener and interview. The unweighted response rate for all beneficiary interviews was 80 percent.
- Differences between weighted and unweighted response rates were small.
- The eligibility rates for organization RPs were lower than those for individual RPs. The reasons-for-ineligibility data indicate that the primary reasons for organization RPs to be ineligible are that they are fee-for-service organizations or that they serve 50 or more beneficiaries.

Table 4-4. Unweighted and weighted eligibility rates and response rates by domain for RP eligibility screener and RP interviews

Domain	n	Eligibility Rate (%)		Response Rate (%) RP Screener		RP Interview	
		Unwtd	Wtd	Unwtd	Wtd	Unwtd	Wtd
All	5,098	92	96	92	91	91	91
Beneficiary Age							
0-17	1,465	94	97	91	91	91	90
18 to 64	2,549	92	94	91	92	90	91
65+	1,084	91	93	96	96	95	95
Type of Benefit							
Title 2	2,613	90	95	94	93	93	92
Title 16	1,965	94	97	90	89	90	89
Both	520	95	97	92	91	90	91
Region							
Northeast	940	91	96	91	89	90	89
Midwest	1,196	91	96	91	91	90	90
South	1,971	94	96	93	92	92	92
West	991	92	96	92	92	91	91
MSA/Non-MSA							
MSA	4,114	92	96	92	91	91	90
Non-MSA	984	92	94	92	93	94	93
RP type, relationship, # beneficiaries							
Indiv., parent	1,501	96	97	91	91	91	91
Indiv., adult child or other relative	1,390	95	95	93	92	93	92
Indiv., unrelated, 1 beneficiary	858	90	92	88	88	86	87
Indiv., unrelated, 2+ beneficiaries	884	91	92	92	92	91	91
Organization	465	77	79	99	99	98	99
Beneficiary's monthly issuance							
$50-$399	1,155	93	96	92	91	91	91
$400-$599	858	92	96	93	92	92	91
$600-$699	1,869	93	97	90	89	89	89
$700+	1,216	90	95	95	94	94	94

Table 4-5. Unweighted and weighted eligibility rates and response rates by domain for beneficiary interview

Domain	n	Eligibility Rate (%)		Interview Response Rate (%)	
		Unweighted	Weighted	Unweighted	Weighted
All	2,543	87	92	80	79
Beneficiary Age Group					
14-17	519	93	96	76	77
18 to 64	1,420	89	92	80	81
65+	604	78	79	83	80
Type of Benefit					
Title 2	1,324	83	90	80	78
Title 16	925	92	94	79	81
Both	294	93	96	81	80
Region					
Northeast	441	88	92	79	77
Midwest	596	85	91	79	75
South	1,020	89	93	82	81
West	486	85	92	79	79
MSA/Non-MSA					
MSA	2,006	87	93	78	76
Non-MSA	537	88	90	88	87
RP type, relationship, # beneficiaries					
Indiv., parent	701	87	95	79	79
Indiv., adult child or other relative	716	86	89	79	78
Indiv., unrelated, 1 beneficiary	451	89	87	77	76
Indiv., unrelated, 2+ beneficiaries	438	89	87	81	82
Organization	237	68	68	92	92
Beneficiary's monthly issuance					
$50-$399	570	89	92	79	79
$400-$599	414	87	93	78	79
$600-$699	905	91	94	80	80
$700+	654	81	88	82	76

Table 4-6. Unweighted status and disposition counts for all RP eligibility screeners

Status-count symbol	Status Description	Unwtd status count	Disposition code	Disposition description	Disposition count
C_{ER}	Eligible respondents	4,334	CO	Complete	4,334
			PC	Partial complete	0
C_{FI}	Ineligible respondents	361	IN	Ineligible	361
C_{U}	Nonrespondents of unknown eligibility	403	IL	Illness	9
			LP	Language problem	4
			LS	Spanish only	2
			MC	Max calls	11
			NL	Unable to locate	277
			RB	Refusal	27
			UA	Unavailable	26
			ON	Other nonresponse	47
All		5,098			

Table 4-7. Unweighted status and disposition counts for all RP interviews

Status count symbol	Status Description	Unwtd status count	Disposition code	Disposition description	Disposition count
C_{ER}	Eligible respondents	4,297	CO	Complete	4,297
			PC	Partial complete	0
C_{EN}	Nonrespondents known to be eligible	37	IL	Illness	3
			LP	Language problem	0
			LS	Spanish only	1
			MC	Max calls	3
			RB	Refusal	10
			ON	Other nonresponse	20
C_{FI}	Ineligible respondents	361	NF*	Not fielded; ineligible per eligibility screener response	361
C_{U}	Nonrespondents of unknown eligibility	403	NF*	Not fielded; eligibility screener nonrespondent	403
All		5,098			

* Status count based on disposition codes for RP interview and RP eligibility screener.

Table 4-8. Unweighted status and disposition counts for all beneficiary interviews

Status code symbol	Status Description	Unwtd status count	Disposition code	Disposition description	Disposition count
C_{ER}	Eligible respondents	1,771	CB	Completed by beneficiary	1,257
			CP	Completed by proxy	514
			PC	Partial complete	0
C_{EN}	Nonrespondents known to be eligible	133	LP	Language problem	0
			MC	Max calls	1
			NF*	Not fielded; refusal to beneficiary capability screener (beneficiary is eligible per RP eligibility screener)	66
			NP	No proxy available	42
			RB	Refusal by beneficiary	8
			UA	Unavailable	2
			ON	Other nonresponse	14
C_{FI}	Ineligible respondents	282	NF*	Not fielded; ineligible per RP or beneficiary screener response	282
C_{U}	Nonrespondents of unknown eligibility	357	NF*	Not fielded; other	457
All		2,543			

* Status count based on disposition codes for beneficiary interview and RP/beneficiary screener.

Table 4-9. Unweighted disposition counts for all beneficiary capability screeners

Disposition code	Disposition description	Disposition count
NV	Beneficiary not visited; proxy needed	385
PS	Beneficiary passed capability screener	326
CN	Beneficiary capability screener not administered	952
FS	Failed beneficiary capability screener	175
DC	Deceased beneficiary	84
IN	Ineligible	198
LP	Language problem	21
MC	Max calls	9
NF	Not fielded	190
ON	Other nonresponse	12
OA	Outside PSU	78
RB	Refusal by beneficiary	34
RP	Refusal by parent/guardian/other	32
LS	Spanish only	7
NL	Unable to locate	25
UA	Unavailable	15
All		2,543

4.9 Nonresponse Bias Study

The weighted response rates for the RP eligibility screener, the RP interview, and the beneficiary interview were 91 percent, 91 percent, and 79 percent, respectively. A nonresponse bias study was conducted for the beneficiary interview by estimating the nonresponse bias for the following estimators, for which administrative data were available for both respondents and nonrespondents:

- Proportion of the beneficiary population receiving only Title 2 benefits;
- Proportion of the beneficiary population receiving only Title 16 benefits; and
- Average monthly benefit amount for the beneficiary population.

The administrative data associated with these estimators were *not* used in the adjustment of the sampling weights for nonresponse described in Chapter 5.

Table 4-10 and 4-11 contain estimated nonresponse biases for the proportion of the beneficiary population receiving only Title 2 and Title 16 benefits, respectively, within

various domains. None of the estimated nonresponse biases were significantly different from zero, except in the domain of beneficiaries age 65 years and older. In this domain, the estimated Title 2 proportion calculated from only respondents was 2.9 percent higher than the estimate calculated from the entire sample, and the estimated Title 16 proportion calculated from only respondents was 3.9 percent lower than the all-sample estimate.

Table 4-12 contains estimated nonresponse biases for the beneficiary population's average monthly benefit amount. None of the estimated biases were significantly different from zero, except in two domains defined by the type of RP and the relationship of the RP to the beneficiary. In the organization-RP domain and also in the domain in which an individual RP served only one unrelated beneficiary, the average monthly benefit amount estimated from only respondents was significantly larger than the all-sample estimate.

Table 4-10. Comparisons of estimated domain proportions receiving only Title 2 benefits

Domain	Domain Estimates (%)				Estimated nonresponse bias (%)	
	Entire sample		Respondents			
All	55.9	(1.9)	55.8	(1.8)	-0.1	(0.9)
Beneficiary Age						
14 to 17	78.1	(2.4)	76.4	(2.7)	-1.8	(0.9)
18 to 64	30.0	(2.0)	30.4	(2.3)	0.4	(1.0)
65+	79.9	(2.1)	82.8	(1.9)	2.9*	(1.0)
Region						
Northeast	57.6	(4.6)	57.0	(6.2)	-0.6	(2.6)
Midwest	60.8	(3.5)	61.9	(3.9)	1.1	(1.6)
South	52.8	(2.7)	52.0	(2.5)	-0.6	(2.6)
West	56.5	(4.6)	57.4	(5.1)	0.9	(1.7)
MSA/Non-MSA						
MSA	57.6	(2.0)	57.5	(2.4)	-0.1	(0.9)
Non-MSA	50.5	(3.9)	50.3	(2.7)	-0.1	(1.5)
RP type, relationship, # beneficiaries						
Indiv., parent	58.4	(2.4)	57.3	(2.6)	-1.1	(0.9)
Indiv., adult child or other relative	53.2	(2.8)	54.5	(2.6)	1.3	(1.4)
Indiv., unrelated, 1 beneficiary	37.8	(2.5)	38.8	(2.5)	1.0	(1.1)
Indiv., unrelated, 2+ beneficiaries	44.9	(2.7)	44.8	(2.6)	-0.1	(1.1)
Organization	62.5	(3.8)	64.7	(4.0)	2.2	(1.6)

*Significantly different from zero ($p<0.05$).

Note: Estimates were calculated from entire beneficiary sample (using base weights) and from only respondents to beneficiary interviews (using adjusted weights). Standard errors are in parentheses.

Table 4-11. Comparisons of estimated domain proportions receiving only Title 16 benefits

Domain	Domain Estimates (%) Entire sample		Respondents		Estimated nonresponse bias (%)	
All	34.4	(1.7)	34.8	(1.7)	-0.4	(0.7)
Beneficiary Age						
14 to 17	19.4	(2.1)	20.8	(2.4)	1.4	(1.0)
18 to 64	53.6	(2.5)	54.3	(2.8)	0.7	(1.0)
65+	11.5	(1.6)	7.7	(1.3)	-3.9*	(1.0)
Region						
Northeast	34.1	(3.1)	34.3	(4.6)	0.2	(2.2)
Midwest	28.7	(3.7)	28.1	(4.2)	-0.6	(1.6)
South	35.3	(2.5)	36.5	(2.5)	0.2	(2.2)
West	39.0	(4.0)	38.0	(4.5)	-0.9	(1.4)
MSA/Non-MSA						
MSA	33.9	(1.7)	34.5	(2.1)	0.5	(0.8)
Non-MSA	35.8	(4.4)	35.6	(4.2)	-0.2	(0.8)
RP type, relationship, # beneficiaries						
Indiv., parent	33.3	(2.1)	34.3	(2.2)	1.1	(0.8)
Indiv., adult child or other relative	35.6	(2.4)	35.0	(2.6)	-0.6	(1.3)
Indiv., unrelated, 1 beneficiary	50.2	(2.5)	50.0	(2.5)	-0.2	(1.2)
Indiv., unrelated, 2+ beneficiaries	42.9	(3.0)	42.7	(3.0)	-0.2	(1.2)
Organization	19.5	(2.4)	17.7	(2.5)	-1.7	(1.1)

*Significantly different from zero ($p<0.05$).

Note: Estimates were calculated from entire beneficiary sample (using base weights) and from only respondents to beneficiary interviews (using adjusted weights). Standard errors are in parentheses.

Table 4-12. Comparisons of estimated domain-average monthly benefit amounts

Domain	Domain Estimates ($) Entire sample		Respondents		Estimated nonresponse bias ($)	
All	569	(9)	569	(11)	0	(4)
Beneficiary Age						
14 to 17	511	(11)	498	(13)	-13	(8)
18 to 64	578	(13)	585	(15)	7	(5)
65+	739	(18)	754	(21)	15	(8)
Region						
Northeast	607	(24)	617	(32)	10	(13)
Midwest	577	(16)	576	(14)	-1	(8)
South	536	(12)	532	(15)	10	(14)
West	610	(29)	612	(26)	3	(10)
MSA/Non-MSA						
MSA	592	(11)	592	(13)	0	(5)
Non-MSA	497	(14)	497	(15)	0	(5)
RP type, relationship, # beneficiaries						
Indiv., parent	521	(9)	516	(11)	-5	(6)
Indiv., adult child or other relative	645	(21)	652	(21)	7	(9)
Indiv., unrelated, 1 beneficiary	623	(12)	638	(13)	15*	(6)
Indiv., unrelated, 2+ beneficiaries	591	(17)	599	(18)	8	(7)
Organization	645	(23)	662	(24)	17*	(7)

*Significantly different from zero ($p<0.05$).

Note: Estimates were calculated from entire beneficiary sample (using base weights) and from only respondents to beneficiary interviews (using adjusted weights). Standard errors are in parentheses.

References

American Association for Public Opinion Research (2006). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys.* Ann Arbor, MI: AAPOR.

5. Weighting and Variance Estimation

This chapter discusses the procedures used to develop analysis weights for the data collected on RPs and beneficiaries. It also describes the variance estimation procedures used, including the development of replicate weights.

5.1 Development of Analysis Weights

To estimate population parameters, analysis weights are applied to sample data to compensate for a variety of factors. The following three types of analysis weights were calculated:

- RP-caseload weights to be applied to RP-provided data to estimate population parameters about beneficiaries who have RPs;
- RP-level weights to be applied to RP-provided data to estimate population parameters about RPs; and
- Beneficiary weights to be applied to data associated with the beneficiary sample to estimate population parameters about beneficiaries age 14 and older who have RPs.

Analysis weights were developed so that weighted sample data would represent the target populations and the intersections of these populations with various analysis domains. Sample weighting was carried out to accomplish the following objectives:

- Compensate for differential probabilities of selection arising from the multiple stages of sampling;
- Reduce biases occurring because nonrespondents may have different characteristics than respondents; and
- Reduce the variances of the estimates by using auxiliary information.

A four-step procedure was used for developing analysis weights. First, base weights that adjusted for the probability of selection were calculated and assigned to all sampled cases. Second, the base weights were adjusted for eligibility-screener nonresponse

within subgroups. Third, the weights were additionally adjusted for extended-interview nonresponse. Fourth, raking procedures were used to adjust sample distributions to known totals of RPs and beneficiaries, computed by tabulating the January 2006 RPS file. These processes are described more fully in the following sections.

5.1.1 Base Weights

The base weight for a sampled case is the reciprocal of its unconditional probability of its being included in the sample. Base weights for RP-caseload weights were calculated from the unconditional selection probabilities, $P(\alpha\beta\delta)$, defined in Section 2.7 , for:

- Second-stage sampled dyads; and
- First-stage sampled dyads that were ineligible for second-stage dyad sampling.

Base weights for RP-level weights were calculated by dividing a sampled dyad's RP-caseload base weight by the number of dyads listed on the RPS having the same RP as the sampled dyad.

Beneficiary base weights were calculated from the unconditional beneficiary-selection probabilities, $P(\alpha\beta\delta\varepsilon)$, also defined in Section 2.7, for:

- All beneficiaries in the beneficiary sample; and
- First-stage sampled dyads that were ineligible for second-stage dyad sampling and had a reference beneficiary age 14 or older.

With two exceptions, a factor of 0.5 was included in beneficiary base weights if, according to the RPS, the beneficiary had two RPs. The reason was that a beneficiary with two RPs had an increased probability of being selected. The two exceptions were the two pairs of sampled dyads in which the two beneficiaries were the same person, but the two RPs were different. In each of these pairs, the dyad with the older effective date for the RP address was designated as ineligible, and the 0.5 factor was not included in the beneficiary base weight associated with either the eligible or the ineligible dyad in the pair.

Fifty-seven beneficiary base weights had a factor of 0.5 included because the beneficiary had two RPs. Of these 57 beneficiary base weights, 32 were for sampled beneficiaries, and 25 were for first-stage sampled dyads that were ineligible for second-stage dyad sampling and had a reference beneficiary age 14 or older. (According to the RPS, 1,420,200 beneficiaries have an RP who lives in one of the 60 sampled PSUs, and 4,106 of these beneficiaries have two RPs.)

5.1.2 Eligibility-Screener-Nonresponse Adjustment

During sample selection, sampled dyads could be determined to be ineligible (based on administrative data) for the reasons listed in Table 2-4 in Chapter 2. During data collection, RPs and beneficiaries who completed an eligibility screener were determined to be either eligible or ineligible according to the criteria discussed in Section 4.3.3. For RPs who did not complete an eligibility screener questionnaire, however, the RP interview questionnaire was also not completed, and the eligibility status of the sampled dyad was unknown. Similarly, for beneficiaries when the eligibility screener was not completed, the eligibility status of the beneficiary was unknown.

The second step of the weighting process, eligibility-screener-nonresponse adjustment, distributed the base weights of sample units of unknown eligibility (caused by eligibility-screener nonresponse) to sample units of known eligibility. The following eligibility-screener-nonresponse adjustment factor was calculated:

$$SCFACTOR = \frac{S_{ER} + S_{EN} + S_{FI} + S_U}{S_{ER} + S_{EN} + S_{FI}},$$

where

S_{ER} = the sum of the base weights for *eligible respondents* (i.e., RPs or beneficiaries who the interviewer had determined were eligible, and who had also completed the interview questionnaire);

S_{EN} = the sum of the base weights for *eligible nonrespondents* (i.e., RPs or beneficiaries who the interviewer had determined were eligible, but who had *not* completed the interview questionnaire);

S_{FI} = the sum of the base weights for *field-determined ineligibles* (i.e., RPs or beneficiaries who the interviewer had determined were *ineligible)*;[6] and

S_U = the sum of the base weights for *nonrespondents of unknown ineligibility* (i.e., RPs or beneficiaries who had completed neither the eligibility screener questionnaire nor the interview questionnaire and hence their eligibility status was unknown).

Weights adjusted for eligibility screener nonresponse were then calculated as follows:

- For sample cases that were determined to be ineligible based on administrative data, the adjusted weight was equal to the base weight;
- For eligible respondents, eligible nonrespondents, and field-determined ineligibles, the adjusted weight was equal to the base weight multiplied by *SCFACTOR*; and
- For nonrespondents of unknown eligibility, the adjusted weight was set equal to zero.

The approach used to adjust weights for eligibility-screener nonresponse assumes that the proportion of field-determined ineligible cases among the cases of unknown eligibility status is approximately the same as that among those of known eligibility status. This assumption is more likely to be true if eligibility-screener respondents and eligibility-screener nonrespondents are similar with respect to why they may be ineligible. For example, a dyad is ineligible if the beneficiary has died, so among dyads having beneficiaries of different ages, the underlying assumption may be seriously violated if eligibility-screener response rates are different for different beneficiary ages.

To increase the likelihood that the underlying assumption for the weight adjustment process was satisfied, screener-nonresponse adjustment cells were created based on Census region, MSA versus non-MSA status, and dyad stratum number; and a separate *SCFACTOR* was calculated in each cell. The goal was to create cells within which RPs or beneficiaries had a similar propensity to be of known eligibility. The categorical search algorithm CHAID (Kass, 1980) was used separately on the RP sample and the beneficiary sample, where the dependent variable was whether or not the sample case was of known eligibility. CHAID divides the data into groups in a stepwise fashion so that the propensities between the cells are as different as possible. Through a series of chi-square tests for equality of distributions, CHAID identifies the most important predictors and splits the data set into categories. Each of these categories is further segmented based on

[6] S_{FI} does not include the base weights for sample cases determined to be ineligible based on administrative data.

other predictors. The merging and splitting continues until no more statistically significant predictors are found or until a user-specified stopping rule is met, which was that each adjustment cell must contain at least 30 sample cases. The number of adjustment cells created for screener-nonresponse adjustment was 10 for RP-caseload weights, 12 for RP-level weights, and 5 for beneficiary weights.

5.1.3 Interview-Nonresponse Adjustment

The third step of the weighting process, interview-nonresponse adjustment, distributed the adjusted weights of eligible nonrespondents to eligible respondents. As in the preceding adjustment step, nonresponse adjustment cells were created based on Census region, MSA versus non-MSA status, and the dyad stratum number; and separate adjustment factors were calculated in each cell. The dependent variable for the CHAID analysis was response status, and only the eligible respondents and the eligible nonrespondents were analyzed to create the adjustment cells. The number of cells created for interview-nonresponse adjustment was 10 for RP-caseload weights, 12 for RP-level weights, and 4 for beneficiary weights.

5.1.4 Raking

The fourth step of the weighting process, raking, calculated final weights as follows:

- For nonrespondents of both known and unknown eligibility, the final weight was set equal to zero; and
- For ineligibles (both those based on administrative data and field-determined ineligibles) and for eligible respondents, the final weight was the result of modifying the nonresponse-adjusted weight so that the final weights aggregated to control totals calculated from the RPS.

During sample selection, all dyads in which the RP lived in one of the 60 sample PSUs were assigned to one of 27 dyad strata, and a stratified sample of dyads was selected. The 27 dyad strata were defined in terms of RPS variables that described the type of RP (individual or organization), the RP's number of beneficiaries, the living arrangement of the beneficiary, and the relationship between the RP and beneficiary. At the time of sample

selection, these same RPS variables were used to define nonoverlapping RPS tabulation cells; and the number of beneficiaries, the number of beneficiaries age 14 or older, and the number of unduplicated RPs were calculated in each of the RPS tabulation cells. The same rules for excluding dyads from the dyad sampling frame were used for excluding dyads from the calculated RPS tabulation totals. For example, dyads having invalid state or county codes were excluded from both the dyad sampling frame and from the calculated RPS tabulation totals. (Rows 1 through 3 of Table 5-1 quantify the number of RPs and beneficiaries excluded from the dyad sampling frame.)

The RPS tabulation totals can be combined to yield RPS totals for each of the 27 dyad strata. These totals could be used to calculate post-stratified weights; however, the sample sizes in some dyad strata are very small and, as a result, estimates calculated from such weights may have large biases. The RPS tabulation totals can also be combined to yield derived RPS totals corresponding to the various analysis domains described in Table 2-2. These derived RPS totals correspond individually to larger sets of sampled cases, but unlike the RPS totals for the 27 dyad strata, the derived RPS totals are aggregations of overlapping sets of RPS tabulation cells. For example, Domain 1 in Table 2-2 (the set of RPs representing beneficiaries age 18 to 64) overlaps Domain 7 in Table 2-2 (the set of individual RPs who are parents of the beneficiary).

We formed two different groupings of derived RPS totals such that derived RPS totals in the same grouping individually aggregated nonoverlapping sets of RPS tabulation cells and collectively aggregated the entire set of RPS tabulation cells. These groupings of derived RPS totals are called raking dimensions, and derived RPS totals within the same grouping are called levels. The following were the two raking dimensions and their associated levels that were used to calculate final weights:

- **Dimension 1:**
  - Beneficiaries who are children;
  - Beneficiaries age 18 to 64; and
  - Beneficiaries age 65+.
- **Dimension 2:**
  - Individual RP is parent of the beneficiary;

- Individual RP is an adult child or another relative of the beneficiary;
- Individual RP is unrelated to the beneficiary and serves only one beneficiary;
- Individual RP is unrelated to the beneficiary and serves 2+ beneficiaries; and
- Organization RP.

For each dimension and level, RPS tabulation totals were appropriately combined to obtain control totals for all beneficiaries and beneficiaries age 14 and older. For unduplicated RPs, only one control total was used—the total number of unduplicated RPs satisfying conditions 1 through 4 described in Section 2.1. Included in these various control totals were counts of both eligible and ineligible units. For example, the control totals for beneficiaries included ineligible beneficiaries receiving less than $50 in monthly benefits. This is why the goal of the third weighting step was to have the final weights for **ineligibles and eligible respondents** sum to the various control totals.

Raking can be thought of as multidimensional post-stratification. First, the weights for ineligibles and eligible respondents are modified by post-stratifying them to the control totals in dimension 1. Then, the modified weights are post-stratified to the control totals in dimension 2. The process is iterated until the control totals for all dimensions are simultaneously satisfied (at least within a specified tolerance).

Raking with two dimensions of control totals was used to calculate final RP-caseload weights and final beneficiary weights. The control totals for all beneficiaries were used to calculate final RP-caseload weights, and control totals for beneficiaries age 14 or older were used to calculate final beneficiary weights. A sampled dyad's final **RP-level weight** was calculated by raking the nonresponse-adjusted RP-level weights to the RPS total number of unduplicated RPs satisfying conditions 1 through 4 described in Section 2.1.

Rows 1 through 3 of Table 5-1 contain counts of the number of beneficiaries, beneficiaries age 14 or older, and unduplicated RPs who were excluded from both sample selection and the calculation of control totals because they did not satisfy target population Condition 1 (RP residing inside the 48 contiguous states) or target population Condition 4 (valid state/county code). Rows 5 through 11 of Table 5-1 contain sums of final weights for ineligibles and responding eligibles. Table 5-2 contains the control totals in raking

dimensions 1 and 2. The sum of the control totals within each dimension equals the sums of the final weights of ineligibles and responding eligibles on row 11 of Table 5-1.

5.2 Development of Replicate Weights for Variance Estimation

Variance-estimation methods called *replication methods* provide a general method of estimating variances for complex sample designs and multistep weighting procedures, like those used in the Survey of Social Security Representative Payees and Beneficiaries. The basic idea behind the replication approach is to select subsamples repeatedly from the whole sample, to calculate the statistics of interest for these subsamples, and then to use the variability among these subsample or replicate statistics to estimate variance of the full sample statistics. The subsamples are called *replicates* and the statistics calculated from these replicates are called *replicate estimates*. Weights that can be applied to data from the whole sample to calculate replicate estimates are called *replicate weights*.

Replicate weights were developed for the Survey of Social Security Representative Payees and Beneficiaries using the JK2 replication method, which is a stratified-jackknife replication method that entails sampling two (or three) PSUs from each stratum. Because only one PSU had been sampled from each noncertainty PSU stratum, the 58 noncertainty PSU strata were collapsed into 29 variance strata, with each variance stratum consisting of two PSU strata. The PSU strata were formed (in 2002) by sorting the entire set of 1,884 PSUs by PSU characteristics such as region, per capita income, and percent Hispanic, and then assigning PSUs that were next to each other in the sorted list to the same PSU stratum. This sorting of the PSUs to create the PSU strata also sorted the resulting PSU strata. Hence, noncertainty strata that were next to each other with respect to the PSU-stratification sort were collapsed together to form variance strata, each containing two sampled PSUs.

Table 5-1. Counts of excluded population units and sums of final weights

Row	Description	Population units		
		All beneficiaries	Beneficiaries age 14 or older	Unduplicated RPs
	Counts of excluded population units:			
1	RP residing in AK, HI, or PR	143,109	95,944	105,144

2	Invalid state/county code[1]	853,935	603,445	694,738
3	SUBTOTAL	997,044	699,389	799,882
4	Type of final weight	RP caseload	Beneficiary	RP level
	Sums of final weights:			
	Ineligibles based on administrative data:			
5	Monthly benefits < $50	697,239	361,205	514,156
6	Missing issuance data	30,859	27,801	28,138
	Disagreement in files about beneficiary's	40,209	25,938	21,392
7	Name			
8	Two sampled RPs with same beneficiary	315	169	317
9	Field-determined ineligibles	190,598	168,273	143,571
10	Responding eligibles	4,602,395	2,839,614	3,534,863
11	SUBTOTAL	5,561,615	3,423,000	4,242,436
12	TOTAL	6,558,659	4,122,389	5,042,318

[1] Approximately 80% of the population units with invalid state/county codes have blank SSA geocodes.

Table 5-2. Control totals for raking dimensions 1 and 2

Dimension	Level	All beneficiaries	Beneficiaries age 14 or older
1	Beneficiaries who are children	3,520,469	1,381,854
	Beneficiaries age 18 to 64	1,637,553	1,637,553
	Beneficiaries age 65+	403,593	403,593
	TOTAL	5,561,615	3,423,000
2	Individual RP is parent of the beneficiary	3,972,743	2,047,260
	Individual RP is adult child or other relative of the beneficiary	1,263,349	1,068,158
	Individual RP is unrelated to the beneficiary and serves only one beneficiary	134,978	127,749
	Individual RP is unrelated to the beneficiary and serves 2+ beneficiaries	71,539	63,575
	Organization RP	119,006	116,258
	TOTAL	5,561,615	3,423,000

The two certainty PSUs each contained 20 sampled SSUs. From a variance-estimation point of view, sampled SSUs in certainty PSUs are like noncertainty PSUs. These SSUs had been originally sampled by first sorting all SSUs in each PSU by ZIP code. That ZIP code sort of the SSUs was used to create variance strata for the SSUs in certainty PSUs. Pairs of SSUs next to each other in the ZIP code sort within each certainty PSU were assigned to the same variance stratum. This created 20 variance strata in the certainty PSUs for use in calculating replicate RP-caseload weights and replicate RP-level weights. Only 19 variance strata were created in the certainty PSUs for use in calculating replicate beneficiary weights, because the smaller beneficiary-sample size and the lower beneficiary

response resulted in one SSU containing no ineligible or responding-eligible sampled beneficiaries. Hence, one of the variance strata created for use in calculating replicate beneficiary weights contained three PSUs.

The total number of variance strata was 29+20=49 for calculating replicate RP-caseload weights and replicate RP-level weights, and it was 29+19=48 for calculating replicate beneficiary weights. The JK2 replication method entails using as many replicates and replicate weights as variance strata. The kth replicate is created by randomly deleting one PSU from the kth variance stratum. The kth set of replicate weight is calculated by recomputing the sampling weights as if the remaining sample (without the deleted PSU) were the actual sample.

The replicate weights were generated in a series of steps that parallel the steps used to compute the full-sample weights. First, replicate base weights were calculated, and then nonresponse adjustments and raking were performed for each set of replicate weights, using the replicate base weights in the computations in place of the original base weights. These calculations generated final replicate weights.

The WesVar program can be used to obtain estimates and standard errors using the calculated weights. Table 5-3 contains the RP-caseload estimates, the RP-level estimates, and their associated standard errors, for the percentage of RPs who answered "Yes" to the interview question on whether SSA told them what their responsibilities as an RP would be. In the organization-RP domain, the standard error for RP-level estimate is much larger than that for RP-caseload estimate. The reason is that RPs were selected with probability proportional to the number of beneficiaries they represented, which results in the mean-per-cluster (i.e., RP) estimator being less precise than the mean-per-element (i.e., beneficiary) estimator when significant variability occurs in the number of beneficiaries per RP, which it does in the organization-RP domain.

Table 5-3. Estimates and standard errors for percentage of RPs answering "Yes" to question on whether SSA told them about their RP responsibilities

Domain	RP-caseload estimator		RP-level estimator	
	Estimate (%)	Standard error (%)	Estimate (%)	Standard error (%)
All	95	(0.5)	94	(0.6)
Individual RPs	95	(0.5)	94	(0.6)
Organization RPs	97	(0.8)	94	(2.3)

The WesVar program can also be used to calculate design effects for estimated proportions. The design effect for an estimated proportion is the ratio of the variance of the estimate under the actual sample design to the variance under simple random sampling of the estimate calculated with equal weights, given the same sample size for the domain of interest. Estimated proportions that have a design effects greater than 1.0 are less precise than the estimates calculated with equal weights from a same-size simple random sample. For both the RP and beneficiary samples, control totals were used for post-stratification. Consequently, survey variables that are highly correlated with the control variables can have design effects less than 1.0. Table 5-4 contains the design effects for selected population-level estimated proportions.

Table 5-4. Design effects for selected population-level estimated proportions

Sample	Type of weights	Question	Design effect
Beneficiary	Beneficiary	Beneficiary age	1.1
		How many RPs have you had in your lifetime?	1.6
		Who do you live with?	2.1
		How well could you mange benefits without RP?	3.0
		How frequently were you and RP in touch with other?	3.0
RP	RP-caseload	Are you related to the reference beneficiary?	0.1
		How many beneficiaries have you had in your lifetime?	3.1
		How satisfied are you with your ability to understand beneficiary needs?	4.4
	RP-level	RP age	3.6
		RP's highest level of education	12.9

The availability of replicate weights permits analysts to determine the standard error not only for estimated proportions but also for complex non-linear statistics. Table 5-5 contains estimates and standard errors for the percentage agreement between RP and beneficiary responses for three different survey questions. The estimated percentage agreement for a particular question is calculated from the cross-tabulation of RP responses by beneficiary responses for the question, using beneficiary weights to calculate weighted frequencies. The numerator of the percentage-agreement statistic is the sum of the weights for cells on the main diagonal of the cross tabulation—that is, cells in which RP responses agree with beneficiary responses. The denominator of the percentage-agreement statistic is the sum of the weights for all cells in the cross tabulation. The estimates of percentage agreement in Table 5-5 were calculated using full-sample weights. The standard errors were calculated from the results using replicate weights to estimate percentage agreement. A conservative test for two percentage-agreement estimates being significantly different ($p<0.05$) is that the absolute difference of the percentage-agreement statistics is greater than 1.96*(s1+s2), where s1 and s2 are the standard errors of the two percentage-agreement statistics. Hence, the percentage agreement estimates in Table 5-5 are significantly different ($p<0.05$).

Table 5-5. Estimates and standard errors for percentages differences t=between RP and beneficiary responses

Question	Estimate (%)	Standard error (%)
RP & beneficiary talk about beneficiary housing needs?	64.3	(2.2)
Beneficiary has other sources of income?	86.4	(1.2)
Beneficiary has been without a place to live?	98.6	(0.4)

Reference

Kass, G.V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics. 29*, 119-127.

6. Data Files and Documentation

This chapter contains a discussion of the development, format, and final disposition of data files and documentation for the Survey of Social Security Representative Payees and Beneficiaries (Survey). The topics include survey data files and documentation, quality assurance files and documentation, and procedures for destroying files after the final delivery of files to the National Academies.

6.1 Survey Data Files and Documentation

Because the survey data files were moved into SAS prior to carrying out weighting and other statistical data processing activities, SAS was used to produce the deliverable data files. Westat delivered one file for the Representative Payee (RP) survey data and one file for the beneficiary survey data. Each file contained one record for each completed interview. The files contained variables corresponding to all information collected during the interviews, in addition to the following:

- Composite variables (i.e., new variables that were derived from variables on which survey information was collected);
- Appropriate full sample weights;
- Replicate weights; and
- Information derived from the sample design that is necessary for users who wish to use Taylor series variance estimates.

Because organizations other than the National Academies may have access to these data files, they are de-identified to prevent any possibility of identifying specific RPs or beneficiaries. No direct identifiers (Social Security numbers, names, addresses, dates of birth) are included in these files. Also, the geographic information provided is limited. For example, while indicating which records are grouped in the same primary sampling unit (PSU) and stratum are necessary to support the use of some variance estimation software, information associating those record groups with specific geographic areas was omitted. In consultation with the National Academies, a separate "link file" containing information

such as Social Security numbers and more detailed geographic information (such as state) could be provided on a restricted basis.

The data files are in a fixed-column (ASCII) format. Negative numbers are used to identify missing values. The value "-1" indicates a question that was not asked due to a skip pattern. Responses of "Refusal," "Don't Know," and "Not Ascertained" are indicated by the values "-7," "-8," and "-9," respectively.

The data file deliveries also include SAS programs that can be used to read the data back into SAS format. These programs convert the negative numbers used to identify missing values ("-1," "-7," "-8," "-9") to SAS special missing values (".S," ".R," ".D," "."). The programs also contain descriptive variable labels, as well as the SAS statements (PROC FORMAT steps and FORMAT statements) to create coding schemes (SAS formats) for categorical variables and to associate these labels with the appropriate variables.

Each file is accompanied by a codebook that provides a detailed description of the variables contained in the data. It includes the variable name, variable label, type of variable (numeric/character), and the location in the fixed-column data file. For numeric variables, it lists the possible values along with format values and record counts. For character variables, it lists the special missing values and the beginning and ending value contained in the data frequency.

6.2 Quality Assurance Files and Documentation

As part of the quality assurance efforts for the project, Westat shared a variety of quality assurance files with National Academies staff on an ongoing basis. This material included files and reports on survey eligibility, finalized screener and interview status codes, interview validation, survey data frequencies and cross-tabulations, and de-identified survey data files. The files were delivered as SAS data sets, Excel spreadsheets, PDF files, or RTF files, depending on the purpose and expected use of the file. Updated files were made available to authorized National Academies staff on a weekly basis using a secure FTP web site.

The files were accompanied on the web site by a user guide that described the contents of each of the files. In addition, the SAS data sets that contained the survey data

were supplemented by files containing the SAS code (PROC FORMAT steps and FORMAT statements) to create coding schemes (SAS formats) for categorical variables and associate these labels with the appropriate variables. Once data collection was completed, Westat delivered final quality assurance files and user guide.

6.3 Procedures for Destroying Files after Final Data Delivery

Upon contract close-out, Westat took all appropriate steps to eliminate inadvertent disclosure of data. Westat developed a schedule for all close-out events in consultation with the National Academies Project Officer. This process included the deletion of all survey data stored online (disk storage), as well as the destruction of all offline electronic storage media such as CDs. Hard-copy material that needs to be retained was transferred to a facility designated by the National Academies Study Director in locked containers via a secure carrier. Transmittal sheets list the containers, along with counts of the items by container. For all hardcopy material that was destroyed, Westat contracted with a secure subcontractor that provides shredding services.

7. Data Provided in Support of Analysis

This chapter contains a description of the data tabulations provided by Westat to support analysis conducted by the National Academies.

7.1 Development of Analysis Plan

The analyses were driven by the objectives of the Survey of Social Security Representative Payee and Beneficiaries (Survey). These objectives helped guide the development of the instruments, as documented in a construct map that links the objectives to the instrument items. Planned analyses are mapped to each of these evaluation objectives, with the remainder in a catchall category.

The objectives and related topics and constructs follow.

A. Assess the extent to which identified Representative Payees (RPs) are performing their duties in accordance with Social Security Administration (SSA) standards for RP conduct.

- Satisfaction of/with RP
- Beneficiary knows RP
- Quality of RP-beneficiary communication
- RP keeps records
- Communication on needs
- RP's involvement in managing expenses
- Allocation of funds

B. Learn whether the RP payment policies are practical and appropriate.

- Perceived need for RP
- General level of RP-beneficiary contact, communication

C. Determine the extent to which the current program is serving the needs of beneficiaries with RPs.

- Needs met
- Involvement of RP in meeting needs
- Satisfaction with needs met
- Mistrust, disagreement on use of funds

D. Identify the types of RPs that have the highest risk of misuse of benefits.

E. Determine how well SSA supports RPs.

F. Determine how knowledgeable RPs are about their roles and responsibilities.

Other important descriptive analyses, not tied to evaluation objectives include:

- Demographic characteristics and beneficiary living situation;
- General savings, sources of income;
- Type of benefit received (Title II, Title XVI);
- Experience with RP program;
- RP-beneficiary dyad characteristics; and
- Potential outside influence on survey responses.

Specific analytic data requests for analyses were submitted to Westat by the National Academies (NA), with input from the National Academies Committee on Social Security Representative Payees (Committee) and Westat. Given the complexity of the process and the demanding schedule, an interim deliverable was prepared using over one-half of the data collected. This allowed National Academies and the Committee to review preliminary findings and refine their data requests for the final deliverable. This also gave Westat the opportunity to fine-tune the table generation and quality control procedures. In addition, it provided an opportunity to compare beneficiary responses provided by the beneficiaries themselves to those provided by proxies, and to develop aggregate variables.

7.2 Recoding Variables and Creating Aggregate Variables

Most of the variables were recoded prior to data analysis. Variables were recoded for the following reasons:

- Analytic ease (e.g., recoding "NO" responses from 2, which is easier to key enter, to 0, which is analytically easier to use);
- Adjusting data that reflected selected skip patterns (e.g., coding some skips as "NEVER" responses to complete frequency scale and minimize missing data); and
- Normalizing data, such as grouping continuous variables or truncating the upper end of variables to make the variable distributions more normal.

Although insufficient time was available to evaluate the reliability of the questionnaires, aggregate variables or scales were created based on similar questions within the interviews. Aggregate measures are typically more reliable than individual item variables as they take into account performance on multiple similar items. In addition, "reducing the data" in this way also can help to make the analysis more manageable by requiring the use of fewer variables. Before using the derived variables in analyses, analysts assessed their psychometric properties; only those variables that achieved generally accepted levels of reliability were used in analyses. The derived variables organized by evaluation objective, are as follows.

- Objective A. Assess the extent to which identified RPs are performing their duties in accordance with SSA standards for RP conduct:
  - RP involvement in paying beneficiary's expenses (RP and beneficiary); and
  - RP communication regarding beneficiary's needs (RP and beneficiary).
- Objective C. Determine the extent to which the current program is serving the needs of beneficiaries with RPs:
  - Beneficiary's needs met – actual and subjective impression (RP, beneficiary, and measures of discrepancy between their ratings).
- Objective D. Identify the types of RPs that have the highest risk of misuse of benefits:
  - Substance Abuse (RP and beneficiary).

- Objective F. Determine how knowledgeable RPs are about their roles and responsibilities:
  - Knowledge of RP responsibilities (RP and beneficiary); and
  - Knowledge of when to report to SSA (RP only).

Scale creation occurred via a multistep process. Component variables were specified *a priori* for the preliminary construction of the scales. Scales were created by summing the component variables, which were recoded in the desired direction and were sometimes weighted to adjust the contribution of the variable relative to others in the scale. The distribution, item-total correlations, and *alpha*, if individual component variables were removed, were reviewed for the aggregate variables. Final variables were divided into categories, based on quartiles, terciles, and substantive decisions, to both normalize the data and to create a structure that is most amenable to calculation using WesVar.

Most of the scales were created as parallel measures based on the RP and beneficiary interviews. That is, the scales used similar items and had the same range of possible values. This allowed comparability of RP and beneficiary reports. In addition, two "discrepancy variables" were created that compare the difference between RP and beneficiary reports of the extent to which the beneficiary's needs were met. Discrepancy scores reflect whether the RP reported more, similar, or fewer of the beneficiary's needs met, relative to the beneficiary's report.

A few aggregate variables were created, but failed to yield acceptable measures of internal consistency, despite considerable revisions. These variables included measures of risk of RP misuse of benefits (RP and beneficiary, beneficiary impairment (RP only), and knowledge of when to report to SSA (RP only). The aggregate variables were excluded from the analyses.

7.3 Editing Proxy Responses

Approximately one-third of beneficiary responses were provided by proxies. During initial data cleaning efforts, analyst noted that some proxies (i.e., **xx**%, n =**xx**) provided a date of birth that was substantially different from SSA records. Demographic data for these errant proxy records were excluded from the data files and the analyses,

based on the assumption that the proxies became confused as to whether the questions pertained to themselves or to the beneficiaries.

Much of the data collected during the beneficiary interview were examined to evaluate the reliability of proxy responses. These data fell into the following categories:

- **Most likely to be problematic.** Many of these are attitudinal measures, which have been discussed as potentially being of greatest concern. Examples of these variables follow.
  - Satisfaction with RP or with SSA help
  - Knowledge of RP responsibilities
  - Whether the beneficiary's needs have been met
  - Disagreement with RP on how funds are used
  - Belief that RP keeps records on SSA expenditures (these items could be treated as fact or belief)
- **Possibly problematic.** If proxies confuse their birthdays with beneficiaries' birthdays (due to forgetting proxy status), proxies could confuse a lot of the following items. This lapse would be more likely to occur toward the end of the interview, where the birthday question is placed, since more time had elapsed since the proxy-specific instructions.
  - Demographics, living situation, substance abuse
  - Relationship to and with the RP
  - Communication with the RP and SSA (as the proxy could have talked with RP on the beneficiary's behalf)
  - How well the beneficiary could manage SS benefits without the RP
  - Unmet needs, savings, and who meets the beneficiary's needs

Cross-tabulations of these items were created to look for different patterns of responses between proxies and beneficiaries. These findings were evaluated while bearing in mind differences expected between beneficiaries who were able to complete the interview themselves and those who required a proxy respondent.

7.4 Frequency Distributions and Bivariate Analyses

Once the clean data were recoded and aggregated and analysts understood the effect on beneficiary data of including proxy respondents, the data were analyzed. Initial data analyses were primarily descriptive. They included frequency distributions (e.g., values, percentages) of individual variables. Frequency distributions were generated using unweighted and weighted data.

In addition, a large number of bivariate analyses were performed. These analyses took the form of cross-tabulations and fell into four major categories:

- Comparison of individual items on the beneficiary interview;
- Comparison of individual items on the RP interview;
- Comparison of similar items on the beneficiary and RP interviews as a test of similarity and agreement between the dyad respondents; and
- Evaluation of hypotheses (discussed in Section 7.5).

Many of the cross-tabulations included beneficiary (or RP) age, type of benefits received (Title II, XVI, or both), whether the RP was living with the beneficiary, relationship of the RP to the beneficiary (parent, adult child, other relative, nonrelative, or organization), and other characteristics.

All of the cross-tabulations used weighted data, incorporating the sampling weights to produce appropriate standard errors. The statistical significance of bivariate associations was assessed with *Rao-Scott 3* (similar to the *Chi-square* statistic, adjusted to account for the complex sample) with a cutoff of $p < .05$.

WesVar was used to analyze the data, as WesVar provides valid variance estimation based on complex sample designs. A proprietary Westat software system (TabGen) was used to generate the tabulations, and SAS was used to generate simple frequencies.

7.5 Hypothesis Testing

A large number of tabulations were created to test 20 hypotheses. Most were tested using cross-tabulations with tests for statistical significance (*Rao-Scott 3*), and a few were tested by examination of frequency distributions. These tabulations incorporated both beneficiary and RP items from the instruments, as well as aggregate variables derived from several items. A large number of the tabulations examined items and aggregate variables in relation to characteristics of the beneficiary, RP, and the RP-beneficiary dyad; such characteristics included beneficiary or RP age, how the RP and beneficiary are related (if at all), where the beneficiary typically lives, whether the beneficiary lives with the RP, and the beneficiary's SSA benefit type.

The hypotheses are as follows:

1. Representative payees are not aware of many of their duties;
2. Many representative payees believe their beneficiary could manage without them;
3. Representative payees have little contact with their beneficiary;
4. Representative payees believe beneficiary needs are not being met in housing, food, medical, clothing, and leisure areas;
5. Representative payees rarely communicate with their beneficiary about critical issues such as money, housing, food, clothing, and medical needs;
6. Representative payees believe they get little support or training from SSA;
7. Many representative payees are merely "pass through" to the beneficiary;
8. Many beneficiaries are not being well served by their representative payee;
9. Many beneficiaries do not know if they need a representative payee or why they have one;
10. Many beneficiaries want a new representative payee;
11. Many beneficiaries do not believe their representative payees are meeting their housing, food, clothing, medical, or leisure needs;
12. Many beneficiaries think their representative payee "steals" some of their money;

13. Many beneficiaries have had multiple representative payees;

14. Many beneficiaries rarely communicate with their representative payee about critical issues such as money, housing, food, clothing, and medical needs;

15. Many beneficiaries speak a different language than their representative payee;

16. Many beneficiaries do not have much interaction/support from SSA;

17. Parents with custody do not need to be monitored every year;

18. The SSA annual accounting report completed by representative payees is of little value;

19. Certain subgroups of representative payees should be sampled more frequently to complete an accounting form; and

20. Finding good payees is difficult.

7.6 Table Production

Given the large number of cross-tabulations that were generated (approximately 2,700 for the interim deliverable and 6,000 for the final deliverable), this process was automated. Due to the use of complex data weights, results with variance estimates were produced using WesVar. TabGen was used to generate the cross-tabulations and *Rao-Scott 3* statistical tests. TabGen was programmed to use information specified to generate attractive, user-friendly tables assembled in Excel workbooks. One table is presented on each page with several tables per worksheet, assembled by category (i.e., the first variable by which others were crossed or the hypothesis number). For ease of use, a table of contents is located on the first worksheet of each workbook and on the first page of each worksheet; tables are also numbered.

Complex quality control (QC) procedures were built into the table generation process. Quality control efforts were conducted by staff with diverse backgrounds (e.g., senior analysts, senior programmers, data managers, and research assistants), and each process was reviewed by at least two staff, one of whom was always a senior staff member. Examples of QC efforts included the following.

- Programming was checked against the hard-copy specifications that drove them, and against the cross-tabulations that were output by TabGen as Excel tables.

- The variable recode file was checked against the instruments, within itself for consistency, against the SAS syntax files, and relative to the output (i.e., SAS frequencies and Excel cross-tabulations).

- Aggregate variables were checked against the hard-copy specifications that drove them (including variables that were recoded specifically for them), against the SAS syntax files, and in the Excel cross-tabulations.

- Weighted and unweighted frequencies from SAS were checked against cross-tabulations that were generated by TabGen as Excel tables.