Contract No.: 0600-03-60129
MPR Reference No.: 8978

MATHEMATICA
Policy Research, Inc.

National Beneficiary Survey Round 1 (Volume 1 of 3): Editing, Coding, Imputation, and Weighting Procedures

December 19, 2008

Frank Potter
Eric Grau
Debra Wright
Matt Sloan

Submitted to:

Social Security Administration
Office of Disability and Income Support Programs
ITC Building
500 E. Street, SW 9th Floor
Washington, DC 20254

Project Officer:
Paul O'Leary

Submitted by:

Mathematica Policy Research, Inc.
600 Maryland Ave. S.W., Suite 550
Washington, DC 20024-2512
Telephone: (202) 484-9220
Facsimile: (202) 863-1763

Project Director:
Debra Wright

This page has been left blank for double-sided copying.

ERRATA

(Updated December 20, 2016)

The SF-8 mental component summary (MCS) and physical component summary (PCS) scores provided in the original National Beneficiary Survey (NBS) data files were calculated incorrectly. The original values excluded an intercept constant needed to scale the scores to general population norms. The intercept constant values are -10.11675 for the MCS, and -9.36839 for the PCS.

Because the intercept constants were not applied, the scores provided in the original data files were too high relative to what they should be on the population-based scale. Thus, if comparing NBS respondents to the general population, NBS respondents would appear healthier than they should. However, within the NBS respondent sample, the scores still appropriately represented greater or lesser mental and physical health according to the design of the SF-8.

The MCS and PCS variables included in the current data files have been corrected and are now valid for comparisons to other populations.

This page has been left blank for double-sided copying.

CONTENTS

CONTENTS *(continued)*

Chapter **Page**

APPENDICES

TABLES

TABLES *(continued)*

ACRONYMS

ADLs:	Activities of Daily Living
CAPI:	Computer-assisted personal interviewing
CATI:	Computer-assisted telephone interviewing
CHAID:	Chi-Squared Automatic Interaction Detector
ENs:	Employment Networks
IADLs:	Instrumental Activities of Daily Living
ICD-9:	International Classification of Diseases - 9th Revision
MPR:	Mathematica Policy Research
MSA:	Metropolitan Statistical Area
NAICS:	North American Industry Classification System
NBS:	National Beneficiary Survey
PMSA:	Primary Metropolitan Statistical Area
PSU:	Primary Sampling Unit
SAS:	Statistical software, formerly Statistical Analysis System (SAS os a registered trademark of SAS Institute, Inc., Cary, NS)
SOC:	Standard Occupational Classification
SPSS:	Statistical Package for the Social Sciences (SPSS is a registered trademark of SPSS, Inc., Chicago, IL)
SSA:	Social Security Administration
SSDI:	Social Security Disability Insurance (Title II of the Social Security Act)
SSI:	Supplemental Security Income (Title XVI of the Social Security Act)
SSU:	Secondary Sampling Unit

STATA:	Statistical software (STATA is a registered trademark of StatCorp LP, College Station, TX)
TTY:	Teletypewriter
TTW:	Ticket to Work
TRS:	Telecommunications Relay Service

I. INTRODUCTION

As part of an evaluation of the Ticket to Work and Self-Sufficiency program (TTW), Mathematica Policy Research (MPR) conducted the first round of the National Beneficiary Survey (NBS) in 2004. The survey, sponsored by the Social Security Administration's (SSA) Office of Disability and Income Security Programs, collected data from a national sample of SSA disability beneficiaries (hereinafter referred to as the Representative Beneficiary Sample) and a sample of Ticket to Work (TTW) participants (hereinafter referred to as the Ticket Participant Sample). MPR collected data using computer-assisted telephone interviewing (CATI) with computer-assisted personal interviewing (CAPI) follow-ups of CATI nonrespondents and those who preferred or needed an in-person interview to accommodate their disabilities.

A voluntary employment program for people with disabilities, TTW was authorized by the Ticket to Work and Work Incentives Improvement Act of 1999. The legislation was designed to create market-driven services to help disability beneficiaries become economically self-sufficient. Under the program, SSA provides disability beneficiaries with a "Ticket" or coupon, that they may use to obtain employment-support services, including vocational rehabilitation, from an approved provider of their choice (called Employment Networks or ENs).[1]

A. NBS SAMPLE DESIGN OVERVIEW

SSA implemented the TTW program in three phases spanning three years, with each phase corresponding to about one-third of the states. The initial NBS survey design called for four national cross-sectional surveys (called rounds) of Ticket-eligible SSA disability beneficiaries—

[1] For more information on the Ticket to Work Program, see "Evaluation of the Ticket to Work Program Initial Report," (Thornton, et al. 2004).

one each in 2003, 2004, 2005, and 2006—and cross-sectional surveys of Ticket participants in each of three groups of states (Phase 1, Phase 2, and Phase 3 states)—defined by the year in which the program was rolled out (Bethel and Stapleton 2002).[2] In addition, the design called for the first TTW participant cohort in each group of Ticket roll-out states to be followed longitudinally until 2006. This design was subsequently revised to accommodate Phase 1 data collection starting in 2004 rather than 2003. In addition, the final round was postponed to address the experiences of TTW participants under the new TTW regulations; implemented in July 2008. The fourth round will include a cross-sectional Representative Beneficiary survey as well as a survey of new Ticket Participants and is planned for 2009. Details of the sample design for round 4 have not yet been determined; in a change from the original design, Ticket participants from previous rounds will not be re-interviewed at round 4. Table I.1 gives the original planned sample sizes for all rounds of data collection. Actual sample sizes and number of completed cases is provided in Chapter III.

Two surveys were fielded in round 1 (2004): the first national survey of all beneficiaries (the Representative Beneficiary Sample) and the first cross-sectional survey of Ticket participants in the Phase 1 states (the Ticket Participant Sample).

[2] The Ticket to Work program, implemented in 2002, was phased in nationwide over three years. In 2002, the first year of the program, SSA distributed Tickets in 13 "Phase 1" states: Arizona, Colorado, Delaware, Florida, Illinois, Iowa, Massachusetts, New York, Oklahoma, Oregon, South Carolina, Vermont, and Wisconsin. In the Phase 2 rollout, from November 2002 through September 2003, SSA distributed Tickets in 20 "Phase 2" states and the District of Columbia: Alaska, Arkansas, Connecticut, Georgia, Indiana, Kansas, Kentucky, Louisiana, Michigan, Mississippi, Missouri, Montana, Nevada, New Hampshire, New Jersey, New Mexico, North Dakota, South Dakota, Tennessee, Virginia, and the District of Columbia. The Phase 3 rollout ran from November 2003 through September 2004; SA distributed Tickets in the remaining 17 "Phase 3" states: Alabama, California, Hawaii, Idaho, Maine, Maryland, Minnesota, Nebraska, North Carolina, Ohio, Pennsylvania, Rhode Island, Texas, Utah, Washington, West Virginia, and Wyoming, as well as in American Samoa, Guam, the Northern Mariana Islands, Puerto Rico, and the Virgin Islands.

TABLE I.1

NATIONAL BENEFICIARY AND TTW PARTICIPANT SAMPLE SIZES

Sample[a]			Year 1	Year 2	Year 3	Year 4	All Years[c]
National Beneficiary Samples			7,200	4,800	2,400	1,500	15,900
Longitudinal TTW Participant Samples	Phase 1 Cohorts	(1)[b]	1,000	922	850	784	3,556
		(2)		1,000			1,000
	Phase 2 Cohorts	(1)		1,000	922	850	2,772
		(2)			1,000		1,000
	Phase 3 Cohorts	(1)			1,000	922	1,922
		(2)				1,000	1,000
	Total		1,000	2,922	3,772	3,556	11,250
Total Sample Size			**8,200**	**7,722**	**6,172**	**5,056**	**27,150**

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

[a] Sample sizes refer to number of completed interviews

[b](1)=TTW participant longitudinal sample and (2)=TTW participant cross-sectional supplement

[c] This column is a tabulation of the number of interviews, not the number of sample members. Longitudinal cases may be included multiple times in these counts, depending upon the number of completed interviews for the sample member in question.

For all survey rounds, the NBS used a multi-stage sampling design with a supplemental single-stage sample for some Ticket participant populations. For the multi-stage design, data from SSA on the counts of eligible beneficiaries in each county were used to form primary sampling units (PSUs), consisting of one or more counties. The sample of all SSA beneficiaries (the Representative Beneficiary Sample) was selected from among beneficiaries residing in these PSUs (or, in the case of two counties with a large number of beneficiaries, from secondary sampling units) using age-defined sampling strata. Separate samples of Ticket participants within each phase in the original sample design were selected from all Ticket participants in these PSUs. The Ticket Participant Sample was divided into three strata (within each phase) according to the types of payment system under which SSA paid a service provider: (1) the traditional vocational rehabilitation payment system, (2) the milestone-outcome payment system, or (3) the

outcome-only payment system.[3] The supplemental single stage sample for some Ticket participant populations was drawn from all Ticket participants, not just those in the PSUs, with stratification based upon payment type and whether the participant was in a PSU or not. The Round 1 User's Guide (Wright et al. 2008) contains more information on the sampling design.

B. NBS OBJECTIVES

The NBS is one of several components of an evaluation of the impact of the TTW relative to the current system, the SSA Vocational Rehabilitation Reimbursement Program, which has been in place since 1981. The evaluation includes process as well as an impact and participation analysis. Along with the NBS, the data sources include SSA administrative records and interviews with program stakeholders. The NBS collects data needed for the TTW evaluation that are not available from SSA administrative data or other sources.

The NBS has five key objectives:

1. To provide critical data on the work-related activities of SSI and SSDI beneficiaries, particularly as they relate to the TTW implementation
2. To collect data on the characteristics and program experiences of beneficiaries who use their Tickets
3. To gather information about beneficiaries who do not use their Tickets, and the reasons they do not
4. To collect data that will allow us to evaluate the employment outcomes of Ticket users and other SSI and SSDI beneficiaries
5. To collect data on service use, barriers to work, and perceptions about TTW and other SSA programs designed to help SSA beneficiaries with disabilities find and keep jobs

[3] ENs may choose to be paid under the traditional payment system or under one of two other payment systems developed specifically for the Ticket program: (a) an outcome-only payment system or (b) a milestone-outcome payment system. Under both new payment systems, SSA will make up to 60 monthly payments to the EN for each assigned beneficiary who is not receiving SSDI or SSI payments because of work or earnings. Under the milestone-outcome payment system, SSA pays smaller monthly payments in the event that the beneficiary leaves cash benefits but will also pay the EN for up to four milestones achieved by a beneficiary.

The survey data will be combined with SSA administrative data to provide critical information on access to jobs, and employment outcomes for disability beneficiaries, including those who participate in the TTW program and those who do not. Though some sections of the NBS target beneficiary activity directly related to TTW, most of the survey captures more general information on SSA beneficiaries including their disabilities, interest in work, use of services, and employment. As a result, SSA and external researchers interested in disability and employment issues can use the survey data for other policymaking and program-planning efforts.

C. ROUND 1 SURVEY OVERVIEW

Sample members in both the Representative Beneficiary Sample and the Ticket Participant Sample received the same survey instrument. The NBS collects data on a wide range of topics including employment, limiting conditions, experience with SSA programs, employment services, health and functional status, health insurance, income, and socio-demographic information. The survey items were developed and initially pre-tested as part of a separate contract held by Westat. Revisions were made by MPR to prepare the instrument for CATI/CAPI programming, and additional minor wording changes were made after pre-testing. More information about the questionnaire can be found in the Round 1 User's Guide (Wright, et al. 2008). The survey instrument is available from SSA or MPR upon request.

Round 1 CATI data collection for both samples began in February 2004. Beginning in May 2004, MPR conducted in-person CAPI interviews with beneficiaries who did not respond to the CATI interview, as well as those who could not be located (and whose names and other information were sent to field interviewers for additional locating), or who requested an in-person interview to facilitate their participation in the survey. The survey instrument was identical in each mode. When possible, the interview was attempted with the sample person. If the sample person was unable to complete either a telephone or in-person interview, a proxy

respondent was sought. Proxy interviews were attempted only when the sample member was unable to complete the survey himself or herself due to his/her disability. To promote response among Hispanic populations, the questionnaire was available in Spanish. For languages other than English and Spanish, interpreters conducted the interviews. A number of additional accommodations were made available for those with hearing and/or speech impairments including teletypewriter (TTY), Telecommunications Relay Service (TRS), amplifiers, and instant messaging technology.

As shown in Table I.2, the NBS round 1 sample comprised 9,064 cases selected for the Representative Beneficiary Sample and 1,466 cases for the Ticket Participant Sample (for a total of 10,530 cases).

TABLE I.2

ROUND 1 SAMPLE SIZES, TARGET COMPLETES, AND ACTUAL COMPLETES

Sampling Strata	Sample Size	Target Completes	Actual Completes
National Beneficiary Sample	9,064	7,200	6,520
Ticket Participant Sample	1,466	1,000	1,083
Total Sample Size	10,530	8,200	7,603

Source: NBS, round 1

The round 1 CATI and CAPI data collection was completed in October 2004. Interviews were completed with 6,520 individuals in the Representative Beneficiary Sample and 1,083 people in the Ticket Participant Sample for a total of 7,603 interviews completed. An additional 458 beneficiaries and 73 Ticket participants were determined to be ineligible to participate in the survey.[4] Across both samples, 6,302 cases were completed by telephone and 1,301 were

[4]Ineligible sample members include those who were deceased, no longer living in the continental United States, incarcerated or institutionalized, and those whose benefit status was pending. For Ticket participants,

completed by CAPI. Proxy interviews were completed for 1,997 sample members. The weighted response rate for the Representative Beneficiary Sample was 77.5 percent. The weighted response rates for the Ticket Participant Sample was 80.9 percent.

D. NBS DATA DOCUMENTATION REPORTS

The following reports make up the complete documentation describing the NBS, the round 1 data collection, and the data files:

- ***Editing, Coding, Imputation, and Weighting Report*** (current report). This report summarizes the editing, coding, imputation, and weighting procedures as well as the development of standard errors for the round 1 NBS. It includes an overview of the variable naming, coding, and construction conventions used in the data files and accompanying codebooks; describes how the initial sampling weights were computed to the final post-stratified analysis weight for both the Representative Beneficiary Sample and the Ticket Participant Sample (and describes the procedures for combining these samples); describes the procedures used to impute missing responses; and discusses procedures that should be used to estimate sampling variances for the NBS.
- ***Cleaning and Identification of Data Problems Report*** (Wright and Barrett 2008). This report describes the data processing procedures performed for round 1 of the NBS. It outlines the data coding and cleaning procedures and describes the data problems identified, their origins, and the corrections implemented to create the final data file. The report describes the data issues by sections of the interview and concludes with a summary of types of problems encountered and general recommendations.
- ***User's Guide for Restricted and Public Use Data Files*** (Wright et al. 2008). This report is designed to provide users with information about the restricted use data file and planned public use file, including construction of the files; weight specification and variance estimation; masking procedures employed in the creation of the Public Use File; and a detailed overview of the questionnaire design, sampling, and NBS data collection. The report also contains information covered in the two reports mentioned above including procedures for data editing, coding of open-ended responses, and variable construction; and a description of the imputation and weighting procedures and development of standard errors for the survey.

(continued)
ineligibles also included sample members who left the program after sampling was completed, or were not contacted by telephone and were not eligible for field follow-up.

In addition the following supplemental materials are available from MPR or SSA upon request:

- ***NBS Questionnaire*.** This document contains all items on the round 1 survey and includes documentation of skip patterns, question universe specifications, text fills, interviewer directives, and consistency and range checks.
- ***NBS Restricted Access and Public Use File Codebooks.*** The codebooks provide extensive documentation for each variable on the file including variable name, label, position, variable type and format, question universe, question text, number of cases eligible to receive each item, constructed variable specifications, and user notes. Frequency distributions and means are also included as appropriate.

In the discussion that follows, we document the editing, coding, imputation, and weighting procedures as well as the development of standard errors for the round 1 NBS. Chapter II is an overview of the variable naming, coding, and construction conventions used in the data files and accompanying codebooks. Chapter III describes how the initial sampling weights were computed to the final post-stratified analysis weight for both the Representative Beneficiary Sample and the Ticket Participant Sample; also described are the procedures for combining these samples. Chapter IV describes the procedures used to impute missing responses for selected questions. Chapter V discusses the procedures that should be used to estimate sampling variances for the NBS. Appendix A lists the open-ended items that were assigned additional categories, as discussed in Chapter II. Industry and occupation codes, also discussed in Chapter II, are listed in Appendices B and C. Detailed parameter estimates and standard errors for the weight adjustment models discussed in Chapter III are presented in Appendix D. Appendix E covers the SUDAAN parameters for the national estimates from the TTW round 1 sample.

II. DATA EDITING AND CODING

Prior to imputation, survey data were edited and coded to create an NBS data file. This chapter provides an overview of the variable naming, coding, and construction conventions used in the data files and accompanying codebooks.

A. DATA EDITING

At the start of data cleaning, a systematic review of the frequency counts of the individual questionnaire items was conducted. We reviewed frequency counts by each questionnaire path to identify possible skip pattern errors. We also reviewed interviewer notes and comments in order to flag and correct individual cases. In consultation with SSA and research analysts, we took the general approach of editing only those cases where there appeared to be an obvious data entry or respondent error. As a result, while we devoted a great deal of time to a meticulous review of individual responses, some suspect values remain on the file. (See Wright and Barrett 2008 for full detail regarding the editing and cleaning procedures.)

For all items with fixed field numeric responses (such as number of weeks, number of jobs, and dollar amounts) we reviewed the upper and lower values that had been assigned by interviewers. While data entry ranges were set in the CATI instrument to prevent improbable responses from being entered, these ranges were intentionally set to accommodate a wide range of values to account for the diversity expected in this population, and so that the interview could continue in most situations. For these reasons, extremely high and low values were set to missing (.D=don't know) if it appeared that a data entry error had been made.

The NBS instrument included several consistency edit checks to flag potential problems during the course of the interview. To minimize respondent burden, however, all consistency edit checks were suppressible. While the interviewer was instructed to probe such responses, the

interview could continue beyond the item if the respondent could not resolve the problem. In the post-interview stage, we manually reviewed remaining consistency problems to determine whether the responses were plausible. After investigating these cases, we corrected them or set them to missing when an obvious error was encountered.

During data processing, we created several constructed variables to combine data across items. For these items, both the survey team and the analysis team reviewed the specifications, several reviewers checked the SAS programming code, and we reviewed all data values for the constructed variables based on the composite variable responses and frequencies.

For open-ended items that are assigned numeric codes, we examined frequencies to ensure that valid values were assigned. For health condition coding, we also examined codes to verify that the same codes were not assigned to both main and secondary conditions. Cases coded incorrectly were recoded according to the original verbatim response.

B. CODING OF OPEN-ENDED AND VERBATIM RESPONSES

The NBS questionnaire includes a number of questions designed to elicit open-ended responses. To make it easier to use the data connected with these responses in an analysis, we grouped the responses and assigned them numeric codes when possible. The methodology used to code each variable depended upon the content of the variable.

1. Coding Open-Ended, Other/Specify, and Field Coded Responses

Three kinds of questions (described below) on the NBS did not have designated response categories; rather, the response to these questions was recorded verbatim:

- ***Open-ended questions*** have no response options specified (such as E43—Why are you no longer receiving services from your employment network?). For these items, interviewers recorded the verbatim response. Using common responses, we developed categories and reviewed them with analysts. Coders then attempted to code the verbatim response into an established category. If the response did not fit into one of those categories, it was coded as "other."

- ***"Other/specify"*** is a response option for questions that have a finite number of possible answers that may not necessarily capture *all* possible responses. A good example is: "Did you do anything else to look for work in the last four weeks that I didn't mention?" For questions of this type, respondents are asked to specify an answer to the question "anything else?" or "anyone else?"
- ***Field-coded responses*** are answers coded by interviewers into a pre-defined response category without reading the categories aloud to the respondent. If none of the response options seem to apply, interviewers select an "other specify" category and type in the response.

As part of data processing, we examined a portion of all verbatim responses in an attempt to uncover dominant themes for each question. Based on this initial review, we developed a list of categories and decision rules for coding verbatim responses to open-ended items. In addition, supplemental response categories were added to some field-coded or other/specify items to facilitate coding if there were enough such responses and they could not be back-coded into pre-existing categories. (A list of all open-ended items assigned additional categories during the coding process appears in Appendix A.) Thus we categorized verbatim responses for quantitative analyses by coding responses that clustered together (for open-ended and "other/specify" responses) or by back-coding responses into existing response options if appropriate (for "field-coded" and "other/specify" items). If during coding, it became apparent that changes to the coding scheme were necessary (for example adding additional categories or clarifying coding decisions), new decision rules were discussed and documented. Verbatim responses were sorted alphabetically by item for coders and could be filtered by coding status so that new decision rules could be easily applied to cases that had been previously coded. When it was impossible to code a response, when responses were invalid, or when they could not be coded into a given category, we assigned a two-digit supplemental code to the response (see Table II.1). The verbatim responses themselves are excluded from the data files. (See Barrett and Wright (2008) for full detail regarding the back-coding procedures.)

TABLE II.1

SUPPLEMENTAL CODES FOR OTHER, SPECIFY CODING

Code	Label	Description
94	Invalid Response	Indicates this response should not be counted as an "other" response but should be deleted.
95	Refused	Used only if verbatim indicates respondent refused to answer the question.
96	Duplicate Response	Indicates the verbatim response has already been selected in a 'code all that apply' item.
98	Don't Know	Used only if the verbatim indicates that the respondent does not know the answer.
99	Not Codeable	Indicates that a code cannot be assigned based on the verbatim response.

Source: NBS, round 1

2. Health Condition Coding

Responses to questions on health conditions required a specific type of open-ended coding. In Section B of the questionnaire, each respondent was asked to cite the main and secondary physical or mental conditions that limit the kind or amount of work or daily activities they can do. Main conditions could be reported at one of four items: B2 (main reason limited), B6 (main reason eligible for benefits), B12 (main reason was eligible for benefits if not currently eligible), and B15 (main reason limited when first started getting disability benefits). The main purpose of items B6, B12, and B15 was to collect information on a health condition from people who reported no limiting conditions in B2. For example, if respondents said that they had no limiting conditions, they were asked if they were currently receiving benefits from Social Security. If they answered "yes," they were asked for the main reason that made them eligible for benefits (B6). If respondents said that they were not currently receiving benefits, they were asked whether they had received disability benefits in the last five years. If they answered "yes," they were asked for the condition that made them eligible for Social Security benefits (B12) or for the

reason that first made them eligible if they no longer had that condition (B15). If respondents said that they had not received disability benefits in the last five years, they were screened out of the survey and coded as ineligible. Each response to B2, B6, B12, and B15 was assigned a value for the three constructs. Although respondents were asked to cite one "main" condition in B2, B6, B12, or B15, many listed more than one. These additional responses were maintained under the main condition variable and coded in the order in which they were recorded.

For each item on a main condition, respondents were also asked to list any other, or secondary, conditions. For example, respondents reporting a main condition at B2 were asked at B4 to list other conditions that limited the kind or amount of work or daily activities they could do. Respondents reporting the main reason they were eligible for disability benefits (at B6) were asked at B8 to list other conditions that made them eligible. Finally, respondents who reported that they were not currently receiving benefits and who reported a main condition at B12 (the condition that made them eligible to receive disability benefits in the last five years) were asked at B14 for other reasons that made them eligible for benefits. Those who reported that their current main condition was not the condition that made them eligible for benefits and who were asked for the main reason they were first limited were also asked if there were any other conditions that limited them when they first started receiving benefits (B17).

Respondents' verbatim responses were coded according to the International Classification of Diseases, 9th revision, Clinical Modification (ICD-9-CM) five-digit coding scheme. The ICD-9 is a classification of morbidity and mortality information that was developed in 1950 to index hospital records by disease for data storage and retrieval. The ICD-9 was available in hard copy for each of the coders. Coders, many of whom had previous medical coding experience, attended an eight-hour training session before coding and were instructed to code to the highest level of specificity possible. Responses that were not specific enough for a five-digit code were coded to

four (subcategory) or three digits (category codes). Responses that were not specific enough for even three- or four-digit ICD-9 codes were coded either as a physical problem (not specified) or to broader categories representing disease groups. (See Table II.2 for a list of the broad categorical and supplementary codes.) In cases in which multiple, distinct conditions were provided by the respondent, all conditions were coded (for instance, three distinct conditions would be recorded and coded as B2_1, B2_2, and B2_3).

TABLE II.2

ICD-9 CATEGORY AND SUPPLEMENTAL CODES

Code	Label	Description of ICD-9 codes	Corresponding ICD-9 codes
00	Other	Other and unspecified infectious and parasitic disease; alcohol dependence syndrome and drug dependence; learning disorders and developmental speech or language disorders; complications of medical care, not elsewhere classified	136.0-136.9, 303.00-304.90, 315.00-315.39, 999.0-999.9
01	Infectious and parasitic diseases	Borne by a bacterium or parasite and viruses that can be passed from one human to another or from an animal/insect to a human including tuberculosis, HIV, other viral diseases, and venereal diseases (excluding other and unspecified infectious and parasitic diseases)	001.0-135, 137.0-139.8, 01
02	Neoplasms	New abnormal growth of tissue, i.e., tumors and cancer, including malignant neoplasms, carcinoma in situ, and neoplasm of uncertain behavior	140.0–239.9, 02
03	Endocrine/nutritional disorders	Thyroid disorders, diabetes, abnormal growth disorders, nutritional disorders, and other metabolic and immunity disorders	240.0–279.9, 03
04	Blood/blood-forming	Diseases of blood cells and spleen	280.0–289.9, 04
05	Mental disorders	Psychoses, neurotic and personality disorders, and other non-psychotic mental disorders including mental retardation (excluding alcohol and drug dependence and learning, developmental, speech, or language disorders)	290.0–302.9, 305.00-314.9, 315.4-319, 05

TABLE II.2 *(continued)*

Code	Label	Description of ICD-9 codes	Corresponding ICD-9 codes
06	Diseases of nervous system	Disorders of brain, spinal cord, central nervous system, peripheral nervous system, and senses including paralytic syndromes, and disorders of eye and ear	320.0-389.9, 06
07	Diseases of circulatory system	Heart disease, disorders of circulation, and diseases of arteries, veins, and capillaries	390-459.9, 07
08	Diseases of respiratory system	Disorders of the nasal, sinus, upper respiratory tract, and lungs including chronic obstructive pulmonary disease	460-519.9, 08
09	Diseases of digestive system	Diseases of the oral cavity, stomach, esophagus, and duodenum	520.0-579.9, 09
10	Diseases of genitourinary system	Diseases of the kidneys, urinary system, genital organs, and breasts	580.0-629.9, 10
11	Complications of pregnancy, child birth, and the puerperium	Complications related to pregnancy or delivery, and complications of the puerperium	630-677, 11
12	Diseases of skin/ subcutaneous tissue	Infections of the skin, inflammatory conditions, and other skin diseases	680.0-709.9, 12
13	Diseases of musculoskeletal system	Muscle, bone, and joint problems including arthropathies, dorsopathies, rheumatism, osteopathies, and acquired musculoskeletal deformities	710.0-739.9, 13
14	Congenital anomalies	Problems arising from abnormal fetal development, including birth defects and genetic abnormalities	740.0-759.9, 14
15	Conditions in the perinatal period	Conditions that have origin in birth period even if disorder emerges later	760.0-779.9, 15
16	Symptoms, signs, and ill-defined conditions	Ill-defined conditions and symptoms; used when no more specific diagnosis can be made	780.01-799.9, 16
17	Injury and poisoning	Problems that result from accidents and injuries including fractures, brain injury, and burns (excluding complications of medical care not elsewhere classified)	800.00–998.9, 17
18	Physical problem, NEC	The condition is physical, but no more specific code can be assigned.	No ICD-9 codes
95	Refused	Verbatim indicates respondent refused to answer the question.	No ICD-9 codes

TABLE II.2 *(continued)*

Code	Label	Description of ICD-9 codes	Corresponding ICD-9 codes
96	Duplicate condition reported	The condition has already been coded for the respondent.	No ICD-9 codes
97	No condition reported	The verbatim does not contain or symptom to condition to code.	No ICD-9 codes
98	Don't know	The respondent reports that he/she does not know the condition.	No ICD-9 codes
99	Uncodeable	A code cannot be assigned based on the verbatim response.	No ICD-9 codes

Source: NBS, round 1

We ensured that responses were coded according to the proper protocols in several ways. First, we did an initial quality assurance check, per coder, for the first several cases that were coded. An additional 10 percent of all coded responses were reviewed by a supervisor, including cases coders flagged for review that they were unable to code or did not know how to code. Approximately 8 percent of all cases were recoded. In the course of this work, additional decision rules were developed to clarify and document coding protocol. These decisions were discussed with coders and posted to ensure that responses were coded consistently and accurately throughout the coding process. As for other open-ended items, when new decision rules were added, previously coded responses were reviewed and re-coded if necessary. After the ICD-9 coding was complete, we processed the health condition variables into a series of constructed variables that grouped health conditions into broad disease groups.

3. Industry and Occupation

Information about a sample member's current employment and employment in 2004 was recorded in Section C (current employment) and Section D (employment in 2004) of the questionnaire. For each job, respondents were asked to record their occupation (C2 and D4) and the type of business or industry (C3 and D5) where they were employed. Verbatim responses t

the occupation items were coded using the Bureau of Labor Statistic's 2000 Standard Occupational Classification (SOC).[5] The SOC is a system for classifying all occupations in the economy, including private, public, and military occupations in which work is performed for pay or profit. Occupations are classified on the basis of work performed, skills, education, training, and credentials. The sample member's occupation was assigned one occupation code. The first two digits of the SOC codes classify the occupation to a major group and the third digit to a minor group. For the NBS, we assigned three-digit SOC codes to identify the major group comprising the occupation and the minor groups within that classification (using the 23 major groups and 96 minor). Appendix B lists the three-digit minor groups classified within major groups.

Verbatim responses to the industry items were coded according to the 2002 North American Industry Classification System (NAICS).[6] The NACIS is an industry classification system that groups establishments into categories on the basis of activities in which those establishments are primarily engaged. The NAICS has a hierarchical coding system through which all economic activity is classified into 20 industry sectors. For the NBS, we coded NAICS industries to three digits: the first two numbers specify the industry sector, and the third number specifies the sub-sector. Appendix C lists the broad industry sectors. Both the SOC and the NAICS coding schemes are used in most federal surveys, thus providing uniformity and comparability across data sources.

MPR developed supplemental codes for responses to questions about occupation and industry that could not be coded to a three-digit SOC or NAICS code (see Table II.3). As we did in the health condition coding, we performed an initial quality assurance check, per coder, for the

[5] See *Standard Occupational Classification Manual, 2000* or http://www.bls.gov/soc/ for more information.

[6] See North American Industry Classification System, 2002 or http://www.naics.com/info.htm for more information.

first several cases that were coded. An additional 10 percent of all coded responses were reviewed by a supervisor, including cases coders flagged for review that they were unable to code or did not know how to code. Approximately 4 percent of all cases were recoded.

TABLE II.3

SUPPLEMENTAL CODES FOR OCCUPATION AND INDUSTRY CODING

Code	Label	Description
94	Sheltered Workshop	Code used if occupation is in sheltered workshop and the occupation cannot be coded from verbatim.
95	Refused	The respondent refuses to give his/her occupation or type of business.
97	No occupation or industry reported	No valid occupation or industry is reported in the verbatim.
98	Don’t know	The respondent reports that he/she does not know the occupation or industry.
99	Uncodeable	A code cannot be assigned based on the verbatim response.

Source: NBS, round 1

III. SAMPLING WEIGHTS

The final analysis weights for the Representative Beneficiary Sample and the Ticket Participant Sample were determined via a three-step process: (1) calculate the initial weights, (2) adjust weights for two phases of nonresponse (location and completion), and (3) poststratification. This chapter describes these computations for both the Representative Beneficiary Sample and the Ticket Participant Sample. Section A summarized the procedures used to compute and adjust the sampling weights and the procedures for creating composite weights. (Composite weights were used to combine the Representative Beneficiary Sample and the Ticket Participant Sample to analyze the full beneficiary population.) Procedures to computing the weights for the Representative Beneficiary Sample are described in detail in Section B. Section C covers the same information for the Ticket Participant Sample. Because there was an error in Section E of the CATI and CAPI program, we computed special weights for the analysis of response to this section. The computation of these weights is described in Section D of this chapter. The final section describes the procedures for variance estimation for this survey.

A. COMPUTING AND ADJUSTING THE SAMPLING WEIGHTS: A SUMMARY

1. Representative Beneficiary Sample

The sampling weights for any survey are computed from the inverse selection probability that incorporates the stages of sampling in the survey. The Representative Beneficiary Sample was selected in two stages: primary sampling units (PSUs) were selected as part of the round 1 sampling activities, and the individuals within the PSUs were selected from a current database of

beneficiaries.[7] We used four age-based strata in each PSU. In particular, beneficiaries were stratified into the following age groups: 18- to 29-year-olds, 30- to 39-year-olds, 40- to 49-year-olds, and 50- to 64–year-olds. Because we used a composite size measure to select the PSUs, we can achieve equal probability samples in the age strata and nearly equal workload in each PSU for the Representative Beneficiary Sample.[8]

For the initial beneficiary sample, we selected more sample members than we expected to need, to account for differential response and eligibility rates in the PSUs and sampling strata. This "augmented" sample was randomly partitioned into subsamples (called waves), where only some of the waves were used to form the actual final sample. We released an initial set of waves and then monitored data collection to identify which PSUs and strata required additional sample. After the sample members in the initial waves were released for the final sample, we were able to limit the number of additional sample members (in subsequent released waves) only to those PSUs and strata requiring them, and so were able to achieve sample sizes that were close to our targets. Controlling the release of sample also allowed us to control the balance between data collection costs and response rates. The initial sampling weights were computed based on the inverse of the selection probability for the augmented sample. Naturally, only a subset of the augmented sample was actually released, so these initial weights were adjusted for the actual sample size. The release-adjusted weights were post-stratified to population totals obtained from the Social Security Administration (SSA).[9]

[7] An intermediate stage of sampling of secondary sampling units (SSUs) was used in two PSUs, but for the sake of simplicity, these are generally treated as equivalent to PSUs in this description. All PSUs and SSUs were selected during the round 1 sampling activities.

[8] The composite size measure was computed from the sum of the products of the sampling fraction for a stratum and the estimated count of beneficiaries in that stratum and PSU (Folsom et al. 1987).

[9] These totals were obtained from a frame file provided by the Social Security Administration (SSA) that contains basic demographics for all SSI and SSDI beneficiaries.

To account for nonresponse, we used two response propensity models (using logistic regression methods) to estimate (1) a propensity score for locating a sample member and (2) a propensity score for response among located sample members. In our experience with this survey, factors associated with the inability to locate a person tend to be different from those associated with cooperation. The unlocated person cannot deliberately avoid or otherwise refuse to cooperate. For instance, that person may have chosen not to list his or her number or may frequently move from one address to another, but he or she has not specifically shown an unwillingness to cooperate with the survey itself. Located nonrespondents may delibertately avoid the interviewer or may be expressing displeasure or hostility toward surveys, in general, or the SSA, in particular.

For the two forms of nonresponse, we used logistic regression propensity modeling to identify factors associated with each dimension of nonresponse. To develop these models, we used information from the SSA data files and geographic information (such as urban/rural or geographic region) as covariates. Using a liberal level of statistical significance (0.3) in forward and backward stepwise regression models, we made an initial attempt to reduce the pool of covariates and interactions. Any covariate or interaction that was clearly unrelated to locating the respondent, or to response propensity, was excluded from the pool.

After this initial step was completed, a series of models was carefully evaluated by comparing the R-squared statistic[10] and other measures of predictive ability. Model-fitting also involved reviewing the statistical significance of the coefficients of the covariates in the model. We then used the specific covariate values for each located person (cooperating person) to

[10] The Generalized Coefficient of Determination (Cox and Snell 1989) is a measure of the adequacy of the model, where higher numbers indicate a greater difference between the likelihood of the model in question and the null model likelihood. The "Max rescaled R-Square" scales this value to have a maximum of 1.

estimate a propensity to be located (to cooperate). The inverse of these propensity scores is used to adjust the weights from the prior step of weight computations.

The final step is a series of post-stratification adjustments to have the weights sum to known totals obtained from SSA on various dimensions (specifically, gender, age grouping, and for beneficiaries only, recipient status[11]). After post-stratification, the weights underwent a quality assurance check to verify the correctness of the methods used to compute the weights at each step.

2. Ticket Participant Samples

The Ticket Participant Samples were selected from the population of TTW participants, which was partitioned based on three different payment types in the TTW payment system (outcome-only, milestone-outcome, and traditional vocational rehabilitation).[12] The initial sampling weights for the Ticket Participant Samples were computed based on the inverse of the selection probability for the participant. As with the Representative Beneficiary Sample, we used the PSUs as the primary source of the sample members and, where possible, selected an initial larger (augmented) sample. For participants using the outcome-only payment system, the PSUs in the initial sampling design did not contain enough participants to support analysis tasks--even with all outcome-only participants in the PSUs selected for the sample. As a result, it was necessary to supplement the sample from the PSUs with Ticket participants who were not part of the initial sample design. The sample members within the initial sample design are referred to as

[11] Disability payments were made in the form of Supplemental Security Income (SSI), Social Security Disability Insurance (SSDI), or both.

[12] ENs may choose to be paid under one of two payment systems: (a) an outcome payment system or (b) an outcome-milestone payment system. Under each payment system, SSA will make up to 60 monthly payments to the EN for each assigned beneficiary who is not receiving DI or SSI payments because of work or earnings. Under the outcome-milestone payment system, SSA pays smaller monthly outcome payments in the event that the beneficiary leaves cash benefits, but will also pay the EN up to four milestones for work achieved by a beneficiary.

the clustered sample, and the supplemental sample, which was randomly selected from the entire population of outcome-only participants in two geographic strata (in the PSUs and not in the PSUs), was referred to as the unclustered sample.[13]

As with the Representative Beneficiary Sample, we first computed the weights for the augmented sample and then adjusted them for the number of sample members that were n the final sample.[14] For the Ticket Participant Samples, the size of the samples restricted the procedures useable for the nonresponse adjustments. Because Ticket participants were general easier to locate due to their participation in the Ticket program, the number of Ticket participants who could not be located was very small. Hence, for Ticket participants, we used a single stage weighting class approach that combined the adjustment for locating the sample member and for response among the located sample members. The final adjustment for the participants weight was a poststratification adjustment to the counts of participants within subgroups defined by age and gender in the sampling frame.

3. Composite Weights

Although the Ticket participant population constitutes a small subset of the beneficiary population, some analyses require a sample with enough individuals both within and outside the Ticket participant population. This can be accomplished by combining the Ticket Participant Sample and Representative Beneficiary Sample and using composite weights to account for the fact that the samples have been combined. When conducting analyses representing the beneficiary population, these weights can be used to make estimates about participants within the

[13] The respondents who were in the in-PSU stratum of the unclustered sample were selected with certainty, appearing in both the clustered and unclustered samples. Hence, these duplicate cases had to be accounted for in the weighting process, discussed later.

[14] For the clustered sample of participants using the Outcomes-Only payment system, all participants in the PSUs were selected and were released for data collection.

beneficiary population. (Analyses limited to the participants subpopulation use weights only from the Ticket Participant Sample.)

a. Usage of Composite Weights in Practice

The composite weight can be applied to any sample value in the combined Ticket Participant Sample and Representative Beneficiary Sample. The composite weights were calculated to account for participants that were in both the Representative Beneficiary Sample frame and the Ticket Participant frame. For respondents in the Representative Beneficiary Sample who were not Ticket participants, the composite weight is exactly the same as the original (nonresponse-adjusted) beneficiary weight, since these respondents were the sole representatives of the population of beneficiaries who were not Phase 1 TTW participants. However, for respondents in the Representative Beneficiary Sample who were Ticket participants, and for respondents in the Ticket Participant Sample, both samples represent the Phase 1 TTW participant population. Hence, the composite weights were determined by adjusting the original weights so that the total sum of composite weights for Ticket participants from both samples together would add up to the total number of people in the Participant sampling frame: 21,477.

b. Conceptual Framework for Composite Weights

To compute a survey estimate, Est(Y), using information from both samples (such as, the proportion who are currently working) one cannot simply combine the two samples without adjusting the weights, because the Representative Beneficiary Sample includes a small number of Ticket participants who represent the TTW Phase 1 participant population. The Ticket participant sample represents this same population. Separate estimates can be computed for Ticket participants from each sample and combined using the equation,

$$Est(Y) = \lambda \; Y(Participant) \; + \; (1 - \lambda \,) \, Y(Beneficiary)$$

where Y(Participant) is the survey estimate from the Ticket Participant Sample, Y(Beneficiary) is the survey estimate among Ticket participants from the Representative Beneficiary Sample, and λ is an arbitrary constant between 0 and 1. For the round 1 data, there were 22 Ticket participants who responded to the survey in the Beneficiary Sample, and 1,083 Ticket participants who responded to the survey in the Ticket Participant Sample. In our example, the estimates to be combined are the proportion of the 22 Ticket participants in the Beneficiary Sample who are currently working, and the proportion of the 1,083 people in the Ticket Participant Sample who are currently working. For the sampling variance, V(Y), the estimate is computed using the equation,

$$V(Y) \; = \; \lambda^{2} \; V(Y(Participant)) \; + \; (1 - \lambda)^{2} \, V(Y(Beneficiary))$$

where V(Y(Participant)) is the sampling variance for the estimate from the Ticket Participant Sample, and V(Y(Beneficiary)) is the sampling variance for the estimate among Ticket participants from the Representative Beneficiary Sample. Any value of λ will result in an unbiased estimate of the survey estimate, but not necessarily an estimate with the minimum sampling variance. A lambda value producing a sampling variance at its minimum value results in the shortest confidence interval and, by implication, the most precise point estimate.

A value of lambda that minimizes the variance can be calculated as,

$$\begin{aligned} \lambda \; &= \; 1/V(Y(Participant) \, / \, [1 \, / \, V(Y(Participant)) \; + 1/ \, V(Y(Beneficiary)] \\ &= \; V(Y(Beneficiary)) \; / \; [V(Y(Participant)) \; + V(Y(Beneficiary))]. \end{aligned}$$

In this case, the minimum variance is,

$$V(Y) = [V(Y(Participant)) * V(Y(Beneficiary))] \, / \, [V(Y(Participant)) + V(Y(Beneficiary))].$$

To compute the combined-sample estimate with minimum variance, survey estimates are derived by first computing the estimates for each sample, computing a value of λ for each pair of estimates and then combining the point and variance estimates. Although this process produces minimum variance estimates, it is computer intensive and results in some inconsistencies among estimates for percentages and proportions because of differing values of λ among levels of categorical variables.

In this study, we used an alternative approach, which is to identify one or more values of lambda, select a single value based on these lambdas, and compute combined-sample weights. We computed 16 lambdas for a set of 16 variables identified by TTW Project Part A contract staff.[15] These variables are presented in Table III.1. The median of these lambdas (0.101) was used as the final lambda value for the composite weights because the median is a measure of central tendency less affected by extreme values. This produced a value of lambda that was close to optimal for most variables—that is, it resulted in composite weights that gave minimum variance. For analytic purposes, the simplicity of having a single combined sample weight (based on a single lambda estimate) outweighed any concerns of suboptimality for a small number of variables. To compute the composite weight for each sample member in the Ticket Participant Sample, we used,

WT(Combined) = λ WT(participant nonresponse adjusted weight).

For respondents in the Representative Beneficiary Sample who were Ticket participants,

WT(Combined) = (1 - λ) WT(beneficiary nonresponse adjusted weight).

[15] The TTW contract was split into two parts, Part A and Part B. Sampling, weighting, and imputation procedures were conducted under Part B of the contract and analyses under Part A of the contract.

TABLE III.1

VARIABLES USED TO DEFINE WEIGHT COMPOSITING FACTOR (LAMBDA)

Variable Name	Variable Description
B24	Is the sample member currently working?
B28	Has the sample member been looking for work in the past 4 weeks?
B30	Did the sample member work at a job or business for pay in 2003?
B37	Do the sample member's goals include getting a job, moving up in a job, or learning new job skills?
B47a	Does the sample member see himself or herself continuing to work for pay in the next year
B47b	Does the sample member see himself or herself working and earning enough to stop receiving disability benefits in the next year?
B47c	Does the sample member see himself or herself continuing to work for pay in the next five years?
B47d	Does the sample member see himself or herself working and earning enough to stop receiving disability benefits in the next five years?
K3	What is the sample member's personal income in the last month, before taxes?
L24	What is the sample member's annual 2003 household income?
C_usesvr2003	Did the sample member use state vocational rehabilitation services in 2003?
C_usesmenh2003	Did the sample member use state mental health services in 2003?
C_useothst2003	Did the sample member use other state provider services in 2003?
C_usepriv2003	Did the sample member use private business services in 2003?
C_useoth2003	Did the sample member use other services in 2003?
C_fedpovertylevel	Was the sample member above the federal poverty level in 2003?

Source: NBS, round 1

As described previously, respondents who were in the Representative Beneficiary Sample who were not Ticket participants had no change to their weight. That is,

$$WT(Combined) = WT(beneficiary\ nonresponse\ adjusted\ weight).$$

After the composite weight was computed, point and variance estimates could be computed directly using any survey data analysis software package.

In the following sections, we will describe in more detail the procedures for nonresponse adjustment of weights, first for the Representative Beneficiary Sample and then for the Ticket Participant Samples.

B. REPRESENTATIVE BENEFICIARY SAMPLE

1. Initial Weights

The initial weights were computed using the inverse of the probability of selection. For the Representative Beneficiary Sample, samples were selected independently in each of four age strata in each geographic unit or PSU.[16] The number of sample members selected in each stratum and PSU for the augmented sample was determined by allocating three times the target sample size across the 84 geographic units (PSUs and secondary sampling units) independently for each stratum.[17] This ensured that plenty of reserve sample units were available in case response or eligibility rates were lower than expected. The augmented sample size for the three younger age strata (18 to 29 years, 30 to 39 years, and 40 to 49 years) was 6,000 sample members (three times the target sample size of 2,000) and for beneficiaries 50 to 64 years, the augmented sample size was 3,600 (again three times the target sample size of 1,200). By using the composite size measure described previously, the initial weights for the full augmented sample of 21,600 sample members were calculated by taking the inverse of the global sampling rate (F_i) for each stratum. The global sampling rates and initial weights are given in Table III.2.

[16] The sample of PSUs contained 79 unique selections. Because of the size of its beneficiary population, the PSU representing Los Angeles (LA) county received two selections. Within the LA PSU, secondary sampling units (SSUs) were formed and four SSUs were selected. In the PSU representing Cook County, IL, (Chicago) SSUs were also formed to decrease travel costs and two SSUs were selected. These six SSUs and the other 77 PSUs (843 units) are treated as PSUs for the beneficiary sample.

[17] An augmented sample that was three times as large as needed was selected to allow for adequate supplemental sample in all PSUs and sampling strata within the PSUs and to account for expected variation in the response and eligibility rates across PSUs and sampling strata.

TABLE III.2

SURVEY POPULATION AS OF JUNE 30, 2003, AUGMENTED SAMPLE SIZES AND INITIAL WEIGHTS BY SAMPLING STRATA IN THE NATIONAL BENEFICIARY SURVEY

Sampling Strata (Ages as of June 30, 2003)	Survey Population[a]	Augmented Sample Size[b]	Global Sampling Rate (F_j)	Initial Sample Weights	Released Sample
Beneficiaries between 18 and 29 years old	952,074	6,000	0.006302	158.7	2,514
Beneficiaries between 30 and 39 years old	1,278,292	6,000	0.004694	213.0	2,516
Beneficiaries between 40 and 49 years old	2,362,583	6,000	0.002540	393.8	2,516
Beneficiaries between 50 and 64 years old	4,827,983	3,600	0.000746	1,341.1	1,518
Total	**9,420,932**	**21,600**			**9,064**

Source: Sample allocation and counts computed by MPR.

[a] The survey population represents all Supplemental Security Income (SSI) and Social Security Disability Insurance (SSDI) beneficiaries.
[b] The Augmented Sample was designed to account for the variability in the eligibility and response rates across 320 cells defined by the 80 PSU selections and 4 sampling strata within each PSU. Having a substantial reserve sample in each cell ensured no further sampling was needed and additional sample could be released quickly.

As described previously, the full sample was randomly partitioned into subsamples called waves that mirrored the characteristics of the full sample. The waves were formed in each of the four sampling strata in the 84 geographic units (a total of 336 combinations of PSU and sampling stratum). At the start of data collection, a preliminary sample was assigned to the data collection effort and additional waves of sample were assigned as needed, based on experience with eligibility and response rates. Within the 336 combinations of PSU and sampling strata, the initial weights were adjusted to account for the number of waves assigned to data collection. The final sample size for the Representative Beneficiary Sample was 9,064 beneficiaries, as shown under "Released Sample" in Table III.2.

2. Nonresponse Adjustments

In virtually all surveys, the sampling weights must be adjusted to compensate for the sample members that cannot be located or who, once located, refuse to respond. First, weighted logistic

regression models were fitted where the binary response was whether the sample member could be located. Models were fitted by first identifying a pool of covariates to work from using stepwise regression (as noted earlier), then carefully comparing candidate models using various measures of goodness of fit and predictive ability. This process was repeated among respondents who were located, where another weighted logistic regression model was fitted. The two levels in the binary response for this model were "respondent" versus "nonrespondent." For the Representative Beneficiary Sample, a sample member was classified as a respondent if the sample member or person responding for the sample member completed the interview (that is, an eligible respondent), or the sample member was determined to be ineligible (that is, an ineligible respondent). Ineligible sample members included persons who were never SSA beneficiaries, were in the military service at the time of the survey, were incarcerated, had moved outside of the United States, or were deceased at the time of the survey.

Using the procedures outlined above, the main factors or attributes affecting our ability to locate and interview the sample member included the personal characteristics of the sample member (race, gender, and age), the type of beneficiary (recipient of SSI, SSDI, or both), identity of the payee with respect to the beneficiary, whether the beneficiary and the applicant for benefits lived in the same location, the number of times the beneficiary moved in the past five years (based on information from the SSA "finder" database), primary disability classification, and geographic characteristics.

a. Coding of Survey Dispositions

The status of each sample member was maintained in the MPR Survey Management System during the survey and a final status code was assigned after the completion of all locating and interviewing efforts on a sample member, or at the end of data collection. For the nonresponse adjustments, we classified the final status codes into four categories:

1. Eligible respondents.

2. Ineligible respondents (sample members who were ineligible after sample selection, including deceased, sample members in the military or incarcerated, sample members living outside of the United States, and other ineligible).

3. Located nonrespondents (including active or passive refusals, language barrier situations, and so on).

4. Unlocated sample members (sample members who could not be located either using central office tracing procedures or in-field searches).

This classification of the final status code allowed us to measure the overall response rate, the completion rate among located sample members, and the location rate among all sample members.

b. Response Rates

The response rate for the Representative Beneficiary Sample quoted in the introduction of this document as 77.5 percent is the **weighted overall completion rate**, given in the first line of Table III.3. This response rate is the weighted count of sample members for whom a completed interview was obtained or who were determined to be ineligible, divided by the weighted sample

count of all sample members.[18] It can be determined by taking the product of the weighted location rate and the weighted cooperation rate, also known as the weighted completion rate among located sample members.

TABLE III.3

WEIGHTED LOCATION AND WEIGHTED RESPONSE RATES FOR REPRESENTATIVE BENEFICIARY SAMPLE BY SELECTED CHARACTERISTICS

	Sample	Located Sample		Response among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
All	9,064	8,269	92.5	6,978	83.8	77.5
SSI Only , SSDI Only or Both SSI and SSDI						
SSI Only	3,894	3,449	89.3	2,938	84.3	75.4
SSDI Only	3,323	3,110	94.2	2,595	83.4	78.6
Both SSI and SSDI	1,847	1,710	93.5	1,445	83.8	78.5
SSI or SSDI						
SSI Only or in Both SSI & SSDI Programs	5,741	5,159	90.8	4,383	84.1	76.4
SSDI Only or in Both SSI & SSDI Programs	5,170	4,820	94.0	4,040	83.5	78.6
Constructed Disability Status						
Deaf	122	108	87.3	74	61.0	53.3
Mental	4,841	4,360	90.6	3,662	82.4	74.7
Physical	3,590	3,347	94.3	2,859	85.1	80.4
Unknown	511	454	90.0	383	84.4	76.1
Medical Improvement Expected						
No	8,667	7,931	92.6	6,707	84.0	77.8
Yes	397	338	89.3	271	77.2	69.2

[18] This response rate is the weighted count of sample members for whom a completed interview was obtained or who were determined to be ineligible divided by the weighted sample count of all sample members (# of completed interviews + # partially completed + # of ineligibles) / (# of cases in the sample). It can be determined by taking the product of the weighted location rate and the weighted cooperation rate, also known as the weighted completion rate among located sample members. This response rate is basically equivalent to the AAPOR standard response rate calculation: RR_{AAPOR} = # of completed interviews / (# of cases in the sample - estimated # of ineligible cases). Ineligible cases are included in the numerator for two reasons: (1) the cases classified as ineligible are part of the original sampling frame (and hence the study population). We obtained complete information to fully classify these cases (i.e., their responses to the eligibility questions in the questionnaire are complete) and therefore classify them as respondents; (2) incorporating the ineligibles in the numerator and denominator of the response rate is essentially equivalent to the definition of a response rate with these cases excluded if the persons with an additional estimation of the number of eligible cases among those with eligibility unknown. By including the ineligible cases in the numerator and denominator, we avoid using this estimation stage and the response rate computation is more clearly explicated.

TABLE III.3 *(continued)*

	Sample	Located Sample		Response among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Beneficiaries Age (4 Categories)						
18-29 Years	2,514	2,278	90.7	1,950	85.7	77.8
30-39 Years	2,516	2,258	89.8	1,900	84.2	75.7
40-49 Years	2,516	2,309	91.9	1,950	84.5	77.7
50-64 Years	1,518	1,424	93.9	1,178	82.9	77.9
Sex						
Male	4,687	4,238	91.6	3,534	82.4	75.6
Female	4,377	4,031	93.4	3,444	85.1	79.6
Hispanicity						
Non Hispanic	8,386	7,657	92.6	6,452	83.7	77.6
Hispanic	678	612	91.4	526	84.8	77.5
Race (Detailed)						
White	5,715	5,287	93.6	4,486	84.6	79.3
Black	2,197	1,943	89.8	1,632	82.5	74.1
Unknown	889	807	92.0	680	82.7	76.2
Asian American, Pacific Islander	183	163	87.1	121	71.9	63.2
North American Indian or Alaskan Native	80	69	93.2	59	89.5	83.5
Race (Not Black/Other)						
Not Black	5,978	5,519	93.4	4,666	84.3	78.8
Black and Unknown	3,086	2,750	90.5	2,312	82.6	74.7
Living Situation						
Living Alone	5,084	4,551	90.5	3,856	83.8	75.9
Living with Others	337	313	92.8	269	86.7	81.0
Living with Parents	68	54	79.1	44	79.7	63.1
In Institution or Unknown	3,575	3,351	94.3	2,809	83.6	78.9
Did the Applicant for Benefits Live In Same Zip Code as Beneficiary?						
No	1,029	881	85.6	719	80.8	69.4
Yes	3,817	3,449	91.4	2,994	86.1	78.7
No Information	4,218	3,939	94.0	3,265	83.0	78.0
Identity of the Payee with Respect to the Beneficiary						
Beneficiary Received Beneficiary Payments Himself or Herself	5,214	4,742	92.7	4,006	84.4	78.3
Payee is a Family Member	2,842	2,612	92.4	2,214	82.5	76.2
Payee is an Institution	719	670	93.3	553	82.6	77.1
Other	289	245	86.9	205	83.1	72.4

TABLE III.3 *(continued)*

	Sample	Located Sample		Response among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
SSA Region						
1 Boston	465	428	93.7	340	79.3	74.5
2 New York	655	591	91.1	457	75.5	69.0
3 Philadelphia	1,231	1,148	94.5	1,002	87.2	82.4
4 Atlanta	1,913	1,742	92.8	1,475	84.3	78.2
5 Chicago	1,777	1,646	94.3	1,408	82.7	78.0
6 Dallas	875	795	92.7	700	89.0	82.6
7 Kansas City	441	415	93.6	363	87.8	82.1
8 Denver	334	302	91.2	258	88.5	80.6
9 San Francisco	1,036	908	88.3	731	79.4	70.3
10 Seattle	337	294	87.6	244	84.8	74.4
Census Region						
Midwest	2,218	2,061	94.2	1,771	83.7	78.9
Northeast	1,450	1,324	92.5	1,059	79.1	73.3
South	3,689	3,380	93.2	2,915	86.3	80.4
West	1,707	1,504	88.7	1,233	82.2	73.1
Census Division						
East North Central	1,669	1,542	94.1	1,321	83.1	78.2
East South Central	790	713	91.8	613	87.7	80.6
Middle Atlantic	985	896	92.0	719	78.9	72.8
Mountain	452	403	89.4	339	87.4	78.0
New England	465	428	93.7	340	79.3	74.5
Pacific	1,255	1,101	88.5	894	80.4	71.4
South Atlantic	2,024	1,872	94.0	1,602	84.5	79.4
West North Central	549	519	94.5	450	85.4	80.7
West South Central	875	795	92.7	700	89.0	82.6
MSA / PMSA Size						
Not an MSA / PMSA	1,832	1,698	94.1	1,490	87.0	81.9
Areas of 1 million or more	3,705	3,333	90.9	2,730	81.5	74.1
Areas of 250,000 to 999,999	2,081	1,902	92.7	1,616	84.3	78.3
Areas of 100,000 to 249,999	1,363	1,263	94.0	1,081	84.7	79.6
Areas of less than 100,000	83	73	94.7	61	81.4	76.9
Rural/Urban Continuum Code						
Central Counties of Metropolitan Areas of 1 Million Population or More	3,281	2,940	90.3	2,400	81.1	73.4
Fringe Counties of Metropolitan Areas of 1 Million Population or More	436	403	94.2	342	85.1	80.2
Counties in Metropolitan Areas of 250,000 to 999,999 Population	2,314	2,111	92.9	1,785	84.2	78.3
Counties in Metropolitan Areas of less than 250,000 Population	1,201	1,117	94.5	961	84.2	79.5
Urban Population of 20,000 or more, adjacent to a Metropolitan Area	182	152	86.0	130	85.7	74.0
Urban Population of 20,000 or more, not adjacent to a Metropolitan Area	135	120	90.1	108	89.4	80.5
Urban Population of 2,500 to 19,999, adjacent to a Metropolitan Area	773	733	96.5	630	84.9	81.9
Urban Population of 2,500 to 19,999, not adjacent to a Metropolitan Area	339	318	94.0	282	86.5	81.2

TABLE III.3 *(continued)*

	Sample	Located Sample		Response among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Completely Rural adjacent to a Metropolitan Area	193	182	95.8	167	90.6	86.8
Completely Rural not adjacent to a Metropolitan Area	210	193	92.6	173	92.5	85.7

Source: NBS, round 1

The weighted rates are used because (1) the sampling rates (therefore the sampling weights) vary substantially across the sampling strata as seen in Table III.2; and (2) the weighted rates better reflect the potential for nonresponse bias. The weighted rates represent the percentage of the full survey population for which we were able to obtain information sufficient either to use in the data analysis or to determine as ineligible for the analysis.

c. Factors Related to Location and Response

In addition to overall response rate information, Table III.3 also provides information for selected factors associated with locating a sample member, and factors associated with response among located sample members. The table includes the unweighted counts of all sample members, counts of located sample members, and counts of the sample members for whom a completed interview was obtained or who was determined to be ineligible. The table also includes the weighted location rate, the weighted completion rate among the located sample members, and the weighted overall completion rate for these factors, which helped inform the decision about the final set of variables used in the nonresponse adjustment models.

The **weighted location rate** is the ratio of the weighted sample count for located sample members to the weighted count of all sample members, given in Table III.3 as 92.5 percent. The **weighted cooperation rate** (the weighted completion rate among located sample members), 83.8 percent in Table III.3, is the weighted count of sample members for whom a completed

interview was obtained, or who were determined to be ineligible, divided by the weighted sample count of all located sample members. Weighted cooperation rates reflect the common survey situation that once a person is located, repeated contact efforts will result in a completed interview.

d. Propensity Models for Weight Adjustments

A commonly used method to compute weight adjustments is to form classes of sample members with similar characteristics and to use the inverse of the class response rate as the adjustment factor in that class. The adjusted weight is the product of the sampling weight and the adjustment factor. The "weighting classes" are formed to ensure sufficient counts in each class to make the adjustment more stable (that is to have a smaller variance). The natural extension to the weighting class procedure is to use logistic regression with the weighting class definitions used as covariates, provided each level of the model covariates have sufficient number of sample members to ensure a stable adjustment. The logistic regression approach also has the ability to include both continuous and categorical variables, and standard statistical tests are available to evaluate the selection of variables for the model. For the location and the cooperation weight adjustments, we used logistic models to estimate the propensity for a sample member to be located and to cooperate. The inverse of the propensity score was used as the adjustment factor. The adjusted weight for each sample case is the product of the initial sampling weight and the adjustment factor.

The models were developed using the main effects described previously, plus selected interactions. To identify candidate independent variables and interactions among these variables for the modeling, we first ran a chi-squared automatic interaction detector (CHAID) analysis in SPSS to find possible significant predictors. CHAID is normally attributed to Kass (1980) and Biggs, et al. (1991), and its application in SPSS is described in Magidson (1993). The CHAID

procedure iteratively segments a data set into mutually exclusive subgroups that share similar characteristics based on their effect on nominal or ordinal dependent variables. It automatically checks all variables in the data set and creates a hierarchy that shows all statistically significant subgroups. The algorithm finds splits in the population, which are as different as possible based on a chi-square statistic. It is a forward stepwise procedure; it finds the most diverse subgrouping, and then each of these subgroups is further split into more diverse sub-subgroups. Sample size limitations are set to avoid generating cells with small counts. It stops when splits no longer are significant; that is, that group is homogeneous with respect to variables not yet used or when the cells contain too few cases. The CHAID procedure results in a tree that identifies the set of variables and interactions among the variables that have an association with the ability to locate a sample member (and the propensity of a located sample member to either respond or be ineligible).

The variables and interactions identified using CHAID were then processed using forward and backward stepwise regression (using SAS Logistic procedure with weights normalized to the sample size) to further refine the candidate variables and interaction terms. [19] After identifying a smaller pool of main effects and interactions for potential inclusion in the final model, a set of models were carefully evaluated to determine the final model. Because the SAS logistic procedure does not incorporate the sampling design, the final selection of the covariates was accomplished using the logistic regression procedure in SUDAAN.

For selecting variables or interactions in a model, we included variables or interactions that have a statistical significance level (alpha level) of 0.30 or lower (instead of the standard 0.05). We used a higher significance level because the purpose of the model was to improve the

[19] Because no automated stepwise procedures are available in SUDAAN, the stepwise procedures described here were performed using SAS.

estimation of the propensity score and not to identify statistically significant factors related to response. In addition, the information sometimes reflected proxy variables for some underlying variable that was both unknown and unmeasured. The variables used as main effects and the interaction in the model are summarized in Table III.4 for locating a sample member and in Table III.5 for cooperation among located sample members. The R-squared is 0.024 for the location model and 0.031 for the cooperation model. These values are similar to those observed for other response propensity modeling efforts using logistic regression with design-based sampling weights.

TABLE III.4

LOCATION LOGISTIC PROPENSITY MODEL: REPRESENTATIVE BENEFICIARY SAMPLE

Factors in the Location Model*
Main Effects
MOVE
RACE
REPREPAYEE
PDZIPSAME
SEX
METRO_1M
Two-factor interactions
MOVE * RACE
MOVE * SEX
MOVE * REPREPAYEE
MOVE * PDZIPSAME
PDZIPSAME * REPREPAYEE
SEX * REPREPAYEE
RACE * METRO_1M
RACE * SEX
MOVE * SEX
Three-factor interactions
MOVE * RACE * SEX
MOVE * SEX * REPREPAYEE
MOVE * REPREPAYEE * PDZIPSAME

* The factors are described in the text.

The primary factors in the location model include:

1. ***MOVE.*** The number of address changes in the past five years; three levels: (1) one or two moves, (2) no moves or old information, and (3) no information available about number of moves.[20]

2. ***RACE.*** Two levels: (1) Hispanic, non-Hispanic black, and other/unknown, and (2) non-Hispanic white, non-Hispanic American Indian, and non-Hispanic Asian-Pacific Islander.[21]

3. ***REPREPAYEE.*** The identity of the payee with respect to the beneficiary; four levels: (1) the beneficiary received benefit payments himself or herself, (2) a family member received payments on behalf of the beneficiary, (3) an institution received payments on behalf of the beneficiary, and (4) other or unknown.

4. ***PDZIPSAME.*** Whether the beneficiary and the applicant for benefits lived in the same zip code; two levels: (1) beneficiary and applicant lived in the same zip code, and (2) beneficiary and applicant lived in different zip codes, or information unknown.

5. ***METRO_1M.*** Whether the beneficiary lived in a metropolitan area with a million or more residents; two levels: (1) beneficiary lived in metropolitan area with a million or more residents; and (2) beneficiary lived does not reside in such an area.

6. ***GENDER (SEX).*** Two levels: (1) female, and (2) male or unknown.

Various interactions among these variables were also included in the model for locating the sample member. The main effects using the variable names listed above, as well as interactions, are provided in Table III.4. An expanded form of Table III.4, showing the specific levels of the interactions shown in Table III.4, along with parameter estimates and their standard errors, is provided in Appendix D.

For the cooperation models, the primary factors include:

1. ***PSU1-PSU6.*** Beneficiary residence; six indicator variables with various groupings of PSUs, as identified by CHAID.

2. ***GENDER (SEX).*** Two levels: (1) female, and (2) male or unknown.

[20]A single respondent (out of 9,064) with four moves was included in the category "no information." Nine respondents with three moves were included in the "no moves or old information" category.

[21] Fifty in this category were subsequently identified as Hispanic.

3. ***SSI_SSDI.*** Beneficiary recipient benefit type; three levels: (1) SSDI Only, (2) SSI Only, and (3) both SSI and SSDI.

4. ***DIG.*** Disability diagnostic classification; two levels: (1) deaf or unknown disability, and (2) mental or physical disability (excluding deafness).

5. ***REPREPAYEE.*** The identity of the payee with respect to the beneficiary; two levels: (1) an institution received payments on behalf of the beneficiary, and (2) all other payees.

6. ***PDZIPSAME.*** Whether the beneficiary and the applicant for benefits lived in the same zip code; two levels: (1) beneficiary and applicant lived in the same zip code, and (2) beneficiary and applicant lived in different zip codes, or information unknown.

Once again, various interactions among these variables were also included in the model for the cooperation of the sample members. The main effects using these variable names, as well as interactions, are provided in Table III.5. An expanded form of Table III.5, showing the specific levels of the interactions shown in Table III.5 along with parameter estimates and their standard errors, is provided in Appendix D.

TABLE III.5

COOPERATION LOGISTIC PROPENSITY MODEL: REPRESENTATIVE BENEFICIARY SAMPLE

Factors in the Cooperation Model*
Main Effects
PSU1-PSU6
SSI_SSDI
REPREPAYEE
PDZIPSAME
GENDER
DIG
Two-factor interactions
PSU3 * SEX
PSU1 * PDZIPSAME
PSU1 * SSI_SSDI
PSU2 * REPREPAYEE
PSU1 * DIG
SSI_SSDI * DIG
SSI_SSDI * ZIP
Three-factor interactions
SSI_SSDI * PSU1 * PDZIPSAME

* The factors are described in the text.

3. Post-Stratification

Post-stratification is the procedure in which the weighted sums of the response-adjusted weights are aligned to known totals external to the survey. This process offers face-validity for reporting population counts and has some statistical benefits. For the Representative Beneficiary Sample, we post-stratified to the 24 population totals obtained from SSA.[22] In particular, the totals were the total number of SSI/SSDI beneficiaries by age (four categories), gender, and recipient status (SSI only, SSDI only and both).

C. TICKET PARTICIPANT SAMPLES

As noted earlier, the Ticket Participant Samples were selected from the population of TTW participants, a subset of all SSI/SSDI beneficiaries, which was partitioned based on different payment types in the TTW payment system (outcome-only, milestone-outcome, and traditional vocational rehabilitation). Ticket participants using the traditional payment system accounted for 84 percent (18,100 of 21,477 participants) at the time the sampling frame was developed. Participants using the milestone-outcome payment system totaled 2,809 participants (13 percent) and participants using the outcome-only payment system totaled only 568 participants (2.6 percent). Because the target sample size for each participant stratum was 333, the initial augmented samples of 1,000 and 935 participants were selected among participants using the traditional and the milestone-outcome payments systems, respectively. As was also noted earlier, the PSUs in the initial sampling design did not contain a sufficient number of outcome-only participants to support analysis tasks. As a result, the clustered sample, consisting of respondents

[22] These totals were obtained from a frame file provided by the Social Security Administration (SSA), giving information on basic demographics for all Supplemental Security Income (SSI) and Social Security Disability Income (SSDI) beneficiaries.

selected within the initial sample design, was supplemented by a sample randomly selected from the entire population of outcome-only participants (this was called the unclustered sample).

The clustered sample was part of the original sample design, so respondents were selected from within PSUs, whereas the unclustered sample included units that may or may not have been in the selected PSUs. The unclustered sample was therefore organized into two strata: in the PSU or not in the PSU, where respondents in the PSUs were selected with certainty. Hence, all respondents who were in the clustered sample also appeared in the in-PSU stratum of the unclustered sample. The weights for these duplicate cases had to be appropriately adjusted to account for a single respondent's appearance in two independent samples. The compositing scheme used to do this is discussed in more detail in the next subsection. In addition, respondents who could not be located by the central office[23] based on sample frame information were treated differently in the clustered and unclustered samples. In the clustered sample, potential respondents who could not be located were sent to the field for further follow-up, so that personal interviews could be attempted. In the unclustered sample, no further attempt was made to locate potential respondents who could not be located by the central office. If a sample member was selected as part of both the clustered and unclustered samples, and was sent to the field for further follow-up and located in the field, the response had to be treated differently between the two samples. For the sample respondent, the value in the clustered sample was recorded according to its final status in the field, whereas the value in the unclustered sample was recorded as "ineligible for field follow-up" and was treated as a respondent. Ticket participants were classified as ineligible in the unclustered sample because sample members with no field follow-up (in the unclustered sample) were not "selected" for field follow-up. These respondents are accounted for by those who had field follow-up in the clustered sample (and had

[23] The "central office" is the MPR Survey Operations Center (SOC).

been "selected" for field follow-up). This process is analogous to the accepted practice of subsampling of nonrespondents for more intensive effort—in this case, we subsampled cases for field follow-up. The final sample sizes for the participants sample are in Table III.6.

TABLE III.6

SURVEY POPULATION AND INITIAL AND FINAL SAMPLE SIZES BY SAMPLING STRATA IN THE PARTICIPANT SURVEY

Sampling Strata (Payment System)	Survey Population	Initial Sample Size	Released Sample
Total	21,477	2,626[a]	1,466
1. Traditional	18,100	1,000	441
2. Milestone-Outcome	2,809	935	455
3. Outcome-Only			
Clustered Sample	568	123	123
Unclustered Sample	568	568	447
In PSUs	123	123	123
Not in PSUs	445	445	324

Source: Sample allocation and counts computed by MPR.

[a] The initial and final sample sizes include 123 participants using the Outcomes Only payment system, who were in the PSUs and therefore, were part of both the clustered and unclustered samples. The unique sample sizes are 2,503 for the initial sample and 1,343 in the final sample size.

For the clustered samples for TTW participants, the sample was allocated across the 79 PSUs with the Los Angeles PSU receiving a double allocation because it had two selections. Because of the smaller population sizes, we used only the full PSUs; we did not use the secondary sampling units (SSUs) in the Los Angeles PSU (four SSUs) and in the Cook County (Chicago) PSU (two SSUs), which were used for the Representative Beneficiary Sample.

1. Initial Weights

The initial weights were computed based on the probability of selection within the PSU of the augmented sample and the probability of selection for the PSU. For the unclustered sample for the outcome-only participants, we computed the initial weights based on the selection probability within the two sampling strata (in one of the PSUs and not in any PSU). Since only a

portion of the augmented sample was actually released for use, the initial weights were then adjusted for the sample actually used in the survey.

2. Nonresponse Adjustment and Post-Stratification

Because of the smaller sample sizes compared to the survey of all beneficiaries, we used only a single nonresponse adjustment (combining location and nonresponse) and the classic weighting class method for the nonresponse adjustments. However, we used a single logistic regression model across the three payment types to help form the weighting classes, with payment type included as a covariate in the model. In addition to payment type, variables used in the model were gender and geographic region (Census Region[24]), as well as interactions between these variables. Parameter estimates and their standard errors for this model are given in Appendix D. Only these variables had sufficient cell sizes across the three payment types to be used as viable covariates in the model. Completion rates overall and for each payment type are given for these two variables in Table III.7 (for participants using either the traditional payment system or the milestone-outcome payment system) and Table III.8 (for participants using the outcome-only payment system). Due to the fact that the nonresponse adjustment was accomplished in a single step (unlike the Representative Beneficiary Sample), no location and cooperation response rates were calculated, just an overall completion rate. Completion rates for a large selection of variables, overall and for each payment type, are presented in Tables III.9 through III.11.

[24] The U.S. Census Bureau divides the U.S. into four "Census Regions" (http://www.census.gov/geo/www/us_regdiv.pdf): Northeast (New England, New York, Pennsylvania, New Jersey); Midwest (Ohio, Indiana, Michigan, Illinois, Iowa, Minnesota, Wisconsin, Missouri, Kansas, Nebraska, South Dakota, North Dakota; West (Hawaii, Alaska, Washington, Oregon, California, Arizona, New Mexico, Nevada, Utah, Colorado, Wyoming, Idaho, Montana); South (the remainder of the country)

TABLE III.7

WEIGHTING CLASSES AND UNWEIGHTED AND WEIGHTED RESPONSE RATES BY PAYMENT TYPE

Payment Type	Gender	Census Region	Sample	Response	Unweighted Completion Rate	Weighted Completion Rate
Milestone-Outcome			455	360	79.1	82.0
	Male		237	177	74.7	76.6
		Northeast	68	49	72.1	74.3
		South	39	34	87.2	88.6
		Midwest	74	54	73.0	75.8
		West	56	40	71.4	73.1
	Female		218	183	83.9	87.3
		Northeast	56	46	82.1	84.3
		South	34	27	79.4	87.0
		Midwest	75	62	82.7	87.1
		West	53	48	90.6	90.9
Traditional			441	355	80.5	80.9
	Male		235	189	80.4	80.7
		Northeast	69	54	78.3	78.8
		South & West	108	92	85.2	85.3
		Midwest	58	43	74.1	74.1
	Female		206	166	80.6	81.3
		Northeast	58	40	69.0	69.7
		South & West	87	73	83.9	85.0
		Midwest	61	53	86.9	87.1

Source: NBS, round 1

TABLE III.8

WEIGHTING CLASSES AND UNWEIGHTED AND WEIGHTED RESPONSE RATES
OUTCOMES ONLY PAYMENT TYPE
CLUSTERED AND UNCLUSTERED SAMPLES

	Gender	Census Region	Total	Response	Unweighted Completion Rate	Weighted Completion Rate
Clustered Sample			123	87	70.7	74.9
	Male		76	55	72.4	77.5
		Northeast & Midwest	31	23	74.2	80.4
		South & West	45	32	71.1	74.9
	Female		47	32	68.1	69.7
Unclustered Sample			447	354	79.2	79.9
	Male		241	185	76.8	77.2
		Northeast	65	44	67.7	67.6
		South	42	35	83.3	84.3
		Midwest	75	61	81.3	81.2
		West	59	45	76.3	77.4
	Female		206	169	82.0	82.9
		Northeast	64	54	84.4	85.4
		South	43	34	79.1	80.1
		Midwest	58	45	77.6	78.1
		West	41	36	87.8	89.3

Source: NBS, round 1.

TABLE III.9

WEIGHTED RESPONSE RATES FOR TICKET PARTICIPANT SAMPLE, MILESTONE-OUTCOME PAYMENT TYPE, BY SELECTED CHARACTERISTICS

	Sample	Located Sample	Response Among Located Sample	Overall Respondents
	Count	Count	Count	Response Rate
All	455	423	360	82.0
SSI Only , SSDI Only or Both SSI and SSDI				
SSI Only	170	157	133	81.0
SSDI Only	187	176	145	80.0
Both SSI and SSDI	98	90	82	87.4
SSI or SSDI				
SSI Only or in Both SSI & SSDI Programs	268	247	215	83.6
SSDI Only or in Both SSI & SSDI Programs	285	266	227	82.5
Constructed Disability Status				
Mental	249	227	194	81.5
Physical	171	164	138	82.9
Deaf/unknown	35	32	28	81.1
Beneficiaries Age (4 Categories)				
18-29 Years	91	82	74	84.2
30-39 Years	116	109	98	88.1
40-49 Years	124	118	94	77.2
50-64 Years	124	114	94	79.3
Sex				
Male	237	221	177	76.6
Female	218	202	183	87.3
Hispanicity				
Non Hispanic	401	374	316	82.0
Hispanic	54	49	44	81.8
Race (Detailed)				
White (not Hispanic)	190	174	147	80.8
Black (not Hispanic)	184	174	148	83.1
Other/unknown	81	75	65	82.8
Living Situation				
Living alone	255	234	203	83.2
Other/unknown	200	189	157	80.6

TABLE III.9 *(continued)*

	Sample	Located Sample	Response Among Located Sample	Overall Respondents
	Count	Count	Count	Response Rate
Did the Applicant for Benefits Live In Same Zip Code as Beneficiary?				
Yes	55	49	41	80.8
No	165	151	133	83.9
Unknown	235	223	186	81.1
Identity of Payee with Respect to the Beneficiary				
Beneficiary received benefit payments himself or herself	369	345	296	82.5
Payee is a family member	67	61	51	80.5
Payee is institution/unknown	19	17	13	74.6
SSA Region				
1 Boston	47	45	34	72.2
2 New York	77	70	61	82.1
4 Atlanta	139	130	107	78.4
5 Chicago	71	65	59	86.0
9 San Francisco	89	83	73	82.0
10 Seattle	17	15	12	70.6
99 Other	15	15	14	93.1
Census Region				
Midwest	73	67	61	87.7
Northeast	124	115	95	78.7
South	149	140	116	81.7
West	109	101	88	81.6
Census Division				
East North Central	71	65	59	86.0
Middle Atlantic	77	70	61	82.1
Mountain	92	86	76	83.9
New England	47	45	34	72.2
South Atlantic	148	139	115	80.9
Other	20	18	15	85.0
Metropolitan				
Metropolitan areas of 1 million or more	380	351	301	79.7
Metropolitan Areas of 250,000 to 999,999	60	57	45	79.3
Less than 250,000/not in MSA/PMSA	15	15	14	93.8
Central counties of metropolitan areas of 1 million population	370	341	294	79.9
Counties in metropolitan areas of 250,000 to 999,999 population	70	67	52	78.4
Other	15	15	14	93.8

Source: NBS, round 1.

TABLE III.10

WEIGHTED RESPONSE RATES FOR TICKET PARTICIPANT SAMPLE, TRADITIONAL PAYMENT TYPE, BY SELECTED CHARACTERISTICS

	Sample	Located Sample	Response Among Located Sample	Overall Respondents
	Count	Count	Count	Response Rate
All	441	428	355	80.9
SSI Only , SSDI Only or Both SSI and SSDI				
SSI Only	135	129	104	77.1
SSDI Only	199	195	164	83.3
Both SSI and SSDI	107	104	87	81.5
SSI or SSDI				
SSI Only or in Both SSI & SSDI Programs	242	233	191	79.1
SSDI Only or in Both SSI & SSDI Programs	306	299	251	82.7
Constructed Disability Status				
Mental	215	209	171	79.5
Physical	188	182	159	85.5
Deaf/unknown	38	37	25	66.8
Beneficiaries Age (4 Categories)				
18-29 Years	97	94	75	77.0
30-39 Years	98	93	78	80.8
40-49 Years	159	155	128	81.0
50-64 Years	87	86	74	85.7
Sex				
Male	235	227	189	80.7
Female	206	201	166	81.3
Hispanicity				
Non Hispanic	409	397	327	80.4
Hispanic	32	31	28	88.7
Race (Detailed)				
White (not Hispanic)	250	244	202	81.5
Black (not Hispanic)	134	128	105	78.2
Other/unknown	57	56	48	85.1
Living Situation				
Living alone	228	219	178	78.1
Unknown	213	209	177	84.0
Did the Applicant for Benefits Live In Same Zip Code as Beneficiary?				
No	103	100	83	80.9
Yes	7	7	6	91.2
No information	331	321	266	80.7

TABLE III.10 *(continued)*

	Sample	Located Sample	Response Among Located Sample	Overall Respondents
	Count	Count	Count	Response Rate
Identity of Payee with Respect to the Beneficiary				
Beneficiary received benefit payments himself or herself	340	329	269	79.7
Payee is a family member	76	74	64	84.6
Institution/Unknown	25	25	22	85.9
SSA Region				
1 Boston	11	10	9	83.2
2 New York	116	113	85	74.2
3 Philadelphia	40	38	31	77.5
4 Atlanta	77	75	63	82.1
5 Chicago	161	158	139	86.4
8 Denver	14	13	11	78.6
99 Other	22	21	17	79.2
Census Region				
Midwest	166	163	143	86.2
Northeast	127	123	94	74.8
South	119	115	96	80.8
West	29	27	22	76.5
Census Division				
East North Central	161	158	139	86.4
Middle Atlantic	116	113	85	74.2
Mountain	23	22	17	74.2
New England	11	10	9	83.2
South Atlantic	117	113	94	80.4
Other	13	12	11	84.6
Metropolitan				
Metropolitan Areas of 1 million or more	196	189	147	75.4
Metropolitan Areas of 250,000 to 999,999	144	141	125	87.0
Less than 250,000/Not in MSA/PMSA	101	98	83	82.4
Central counties of metropolitan areas of 1 million population or more	191	184	142	74.8
Counties in metropolitan areas of 250,000 to 999,999 population	149	146	130	87.4
Other	101	98	83	82.4

Source: NBS, round 1.

TABLE III.11

WEIGHTED RESPONSE RATES FOR TICKET PARTICIPANT SAMPLE, OUTCOME-ONLY PAYMENT TYPE, BY SELECTED CHARACTERISTICS

	Sample	Located Sample	Response Among Located Sample	Overall Respondents
	Count	Count	Count	Response Rate
All Outcome-Only Clustered	123	111	87	74.9
SSI Only , SSDI Only or Both SSI and SSDI				
SSI Only	31	29	23	77.1
SSDI Only	75	67	55	78.8
Both SSI and SSDI	17	15	9	56.6
SSI or SSDI				
SSI Only or in Both SSI & SSDI Programs	48	44	32	68.7
SSDI Only or in Both SSI & SSDI Programs	92	82	64	74.2
Constructed Disability Status				
Mental	68	62	47	72.7
Physical	45	40	33	78.6
Deaf/unknown	10	9	7	72.9
Beneficiaries Age (4 Categories)				
18-29 Years	13	13	11	83.2
30-39 Years	32	28	23	70.5
40-49 Years	38	36	29	80.2
50-64 Years	40	34	24	71.0
Sex				
Male	76	69	55	77.5
Female	47	42	32	69.7
Hispanicity				
Non Hispanic	113	101	79	74.2
Hispanic	10	10	8	85.6
Race (Detailed)				
White (not Hispanic)	72	64	49	74.6
Black (not Hispanic)	34	31	24	67.5
Other/unknown	17	16	14	85.6
Living Situation				
Living alone	47	44	32	71.1
Unknown	76	67	55	77.2
Did the Applicant for Benefits Live In Same Zip Code as Beneficiary?				
No	6	5	4	79.3
Yes	35	33	25	70.9
No information	82	73	58	76.6

TABLE III.11 *(continued)*

	Sample	Located Sample	Response Among Located Sample	Overall Respondents
	Count	Count	Count	Response Rate
Identity of Payee with Respect to the Beneficiary				
Beneficiary received benefit payments himself or herself	108	98	77	74.8
Payee is a family member	11	9	6	63.7
Institution/Unknown	4	4	4	100.0
SSA Region				
1 Boston	13	11	11	85.1
2 New York	16	16	8	67.3
4 Atlanta	21	21	16	78.1
5 Chicago	28	24	20	71.4
10 Seattle	26	22	18	69.2
99 Other	19	17	14	86.8
Census Region				
Midwest	28	24	20	71.4
Northeast	29	27	19	74.7
South	21	21	16	78.1
West	45	39	32	74.2
Census Division				
East North Central	28	24	20	71.4
Middle Atlantic	16	16	8	67.3
Mountain	19	17	14	86.8
New England	13	11	11	85.1
South Atlantic	21	21	16	78.1
Other	26	22	18	69.2
Metropolitan				
Metropolitan areas of 1 million or more	115	103	80	71.5
Metropolitan areas of 250,000 to 999,999	8	8	7	90.5
Central counties of metropolitan areas of 1 million population	113	101	78	70.8
Counties in metropolitan areas of 250,000 to 999,999 population	10	10	9	91.5
All Outcome-Only Unclustered	447	394	354	79.9
SSI Only , SSDI Only or Both SSI and SSDI				
SSI Only	94	81	73	78.0
SSDI Only	287	259	233	81.8
Both SSI and SSDI	66	54	48	74.2
SSI or SSDI				
SSI Only or in Both SSI & SSDI Programs	160	135	121	76.4
SSDI Only or in Both SSI & SSDI Programs	353	313	281	80.4
Constructed Disability Status				
Mental	223	197	172	77.8
Physical	185	165	153	83.4
Deaf/Unknown	39	32	29	74.7

TABLE III.11 *(continued)*

	Sample	Located Sample	Response Among Located Sample	Overall Respondents
	Count	Count	Count	Response Rate
Beneficiaries Age (4 Categories)				
18-29 Years	43	38	37	86.2
30-39 Years	122	105	92	75.7
40-49 Years	147	127	118	80.6
50-64 Years	135	124	107	80.9
Sex				
Male	241	211	185	77.2
Female	206	183	169	82.9
Hispanicity				
Non Hispanic	417	366	328	79.3
Hispanic	30	28	26	87.3
Race (Detailed)				
White (not Hispanic)	314	275	249	80.0
Black (not Hispanic)	69	61	50	72.7
Other/Unknown	64	58	55	86.2
Living Situation				
Living alone	153	132	117	77.2
Unknown	294	262	237	81.2
Did the Applicant for Benefits Live In Same Zip Code as Beneficiary?				
No	23	18	18	79.1
Yes	118	102	91	77.6
No information	306	274	245	80.8
Identity of Payee with Respect to the Beneficiary				
Beneficiary received benefit payments himself or herself	392	345	315	81.1
Payee is a family member	37	32	23	62.8
Institution/Unknown	18	17	16	88.2
SSA Region				
1 Boston	86	76	71	82.5
2 New York	42	39	26	63.3
4 Atlanta	122	106	99	81.4
5 Chicago	70	60	56	81.0
8 Denver	19	17	16	83.8
9 San Francisco	39	33	32	83.6
10 Seattle	43	39	34	81.0
99 Other	26	24	20	76.9
Census Region				
Midwest	85	75	69	82.1
Northeast	129	116	98	76.6
South	133	115	106	79.9
West	100	88	81	82.3

TABLE III.11 *(continued)*

	Sample	Located Sample	Response Among Located Sample	Overall Respondents
	Count	Count	Count	Response Rate
Census Division				
East North Central	70	60	56	81.0
Middle Atlantic	42	39	26	63.3
Mountain	57	49	47	83.3
New England	87	77	72	82.7
Pacific	43	39	34	81.0
South Atlantic	126	109	101	80.3
Other	22	21	18	81.8
Metropolitan				
Not in MSA/PMSA	92	80	77	83.7
Metropolitan areas of 1 million or more	219	194	165	76.3
Metropolitan areas of 250,000 to 999,999	83	75	69	83.0
Metropolitan areas of 100,000 to 249,999	53	45	43	81.1
Central and fringe counties of metropolitan areas of 1 million population or more	217	193	164	76.6
Central counties of metropolitan areas of 1 million population	88	79	73	82.7
Counties in metropolitan areas of less than 250,000 population	50	42	40	80.0
Urban population of 20,000 or more	18	17	15	83.3
Urban population of 2,500 to 19,999 adjacent to a metropolitan area	24	17	22	91.7
Urban population of 2,500 to 19,999 not adjacent to a metropolitan area	40	37	32	80.0
Completely rural (no places with a population of 2,500 or more)	10	9	8	80.0

Source: NBS, round 1.

After the nonresponse adjustment, the weights were poststratified to the population totals as obtained from the SSA sampling frame of all SSI or SSDI beneficiaries.

3. Dual Frame Estimation

To obtain estimates for the outcome-only Ticket Participant Samples, it was necessary to combine the clustered and unclustered samples using a "paired sample design." As noted earlier, if a potential respondent in the unclustered sample could not be located in the central office, he or she was considered "ineligible for field follow-up" and nothing further was attempted on that case. However, if a potential respondent in the clustered sample could not be located in the

central office, the case was sent to the field for additional locating efforts (field follow-up). The paired sample design is the methodology used to combine the samples while accounting for these different rules of field follow-up.

Consider the population of participants in the relevant payment type as the target population, which can be separated into two parts: the portion that requires field follow-up and the portion that does not. For the portion of the target population that does not require field follow-up (that is, it can be located by central office locating efforts), both the clustered and unclustered samples are independent samples that can provide unbiased estimates for this subpopulation. However, for the other portion of the target population that does require field follow-up (that is, cannot be located by central office locating efforts), since unclustered sample cases were not eligible for field follow-up, only the clustered sample can provide unbiased estimates for this subpopulation.

Consider first the subpopulation that can be located by central office locating efforts. For this subpopulation, the clustered and unclustered samples can be combined using the compositing methods described previously (called a "dual frame" estimation procedure). More specifically, we used a set of 16 variables to compute the 16 lambda values for the composite weights and used the median value of the lambdas as the final lambda value for the composite weights. The median was used to find a value of lambda that was close to optimal for most variables, and to avoid the effect of very large or very small λ on the measure of a "typical" λ. To compute the composite weight for each sample member in the unclustered sample,

WT(Outcomes) = λ WT(unclustered nonresponse adjusted weight).

For units in the clustered sample,

WT(Outcomes) = (1 - λ) WT(clustered nonresponse adjusted weight).

These weights were for estimates of persons who could be located by central office locating efforts.

Next, consider the subpopulation of persons who could be not located by central office locating efforts. In the unclustered sample, persons who could not be located by central office tracing were considered ineligible, so only the clustered sample could be used. Hence, in this case, no combining is required, and the clustered weight is used directly:

WT(Outcomes) = 1 * WT(clustered nonresponse adjusted weight).

Since the weights for each subpopulation sum up to the total number of individuals in each subpopulation, the two subpopulations can simply be combined to form the entire target population.

These weights were poststratified to the population total (568 participants) for participants using the outcome-only payment system.

D. SECTION E WEIGHTS

There were two problems in Section E of the CATI questionnaire that resulted in eligible respondents not being asked certain questions. For example, all SSI beneficiaries were to be asked question E3, SSI beneficiaries who were 25 or younger at time of sampling (September 2004) and received SSI benefits before age 22 were to be asked question E12, and all SSDI beneficiaries were to be asked question E15. An instrument programming error resulted in beneficiaries who received both SSI and SSDI being skipped out of these questions. Additionally, if the interview of a sample member was interrupted and subsequently continued, the information about whether the case was SSI or SSDI was not kept in the system. Both problems were quickly fixed when discovered, but a number of respondents were not asked these questions (see Table III.12). These three questions are:

- ***E3 asks the respondent*** if the beneficiary (who might be the respondent) has ever heard of a Plan for Achieving Self-Support or a PASS Plan. This is a Social Security incentive that lets beneficiaries set aside money to be used to help them reach a work goal. The money set aside does not affect their benefits.
- ***E12 asks the respondent*** if the beneficiary (who might be the respondent) has ever heard of the student earned-income exclusion. This is a Social Security incentive where if a beneficiary is in school, up to $1,340 of earnings per month are not counted when Social Security figures the benefit.
- ***E15 asks the respondent*** if the beneficiary (who might be the respondent) has ever heard of a Trial Work Period. This is a Social Security incentive that lets beneficiaries earn above $800 per month for nine months without losing their benefits.

TABLE III.12

QUESTIONS E3, E12 AND E15 SKIP PATTERNS

	Question		
	E3	E12	E15
Total	7,603	7,603	7,603
Asked	1,796	440	1,898
Incorrectly Skipped	2,762	491	2,728
Logically skipped	3,045	6,672	2,977

Source: NBS, round 1.

More details about Section E of the instrument can be found in the round 1 User's Guide (Wright, et al. 2008).

To analyze Section E data, a revised set of weights was calculated that accounted for the missing data. In particular, the weights for Section E analyses were computed by ratio-adjusting the weights for beneficiaries and participants who did answer the questions to account for those beneficiaries and participants who were eligible to be asked the questions but were skipped. Table III.13 summarizes the count of eligible sample members and the estimated eligible population size for each of three questions (E3, E12, and E15) and the number and weighted estimate of the population to whom the questions were asked correctly. Table III.14 provides the weighted proportion of the eligible population who were asked the appropriate question.

TABLE III.13

SUMMARY OF RESPONSE TO QUESTIONS E3, E12 AND E15

	Combined Samples		Beneficiaries		Participants	
	Count	Estimated Population	Count	Estimated Population	Count	Estimated Population
Eligible for Question E3 (All SSI Beneficiaries)						
Total	6,022	4,310,642	5,151	4,314,173	871	11,371
SSI Only	2,977	2,804,382	2,664	2,805,544	313	6,288
Both SSI and SSDI	1,581	1,506,259	1,369	1,508,629	212	5,083
Asked Question E3						
Total	2,762	2,620,087	2,459	2,652,712	303	7,271
SSI Only	1,439	1,325,337	1,297	1,327,375	142	3,146
Both SSI and SSDI	1,323	1,294,750	1,162	1,325,337	161	4,125
Eligible for Question E12 (SSI Beneficiary Younger than 25 at Sample Selection and Who Received Benefits before age 22)						
Total	931	422,908	865	424,170	66	1,934
SSI Only	745	342,557	699	343,562	46	1,321
Both SSI and SSDI	186	80,351	166	80,608	20	613
Asked Question E12						
Total	440	200,991	410	200,422	30	797
SSI Only	397	181,789	370	181,257	27	741
Both SSI and SSDI	43	19,202	40	19,165	3	57
Eligible for Question E15 (All SSDI beneficiaries)						
Total	4,626	5,982,441	3,856	5,981,279	770	14,819
SSDI Only	3,045	4,476,181	2,487	4,472,650	558	9,736
Both SSI and SSDI	1,581	1,506,259	1,369	1,508,629	212	5,083
Asked Question 15						
Total	1,898	2,694,883	1,563	2,692,870	335	5,984
SSDI Only	1,641	2,483,418	1,356	2,481,771	285	5,077
Both SSI and SSDI	257	211,466	207	211,099	50	907

Source: NBS, round 1.

TABLE III.14

SUMMARY OF RESPONSE ADJUSTMENTS FOR QUESTIONS E3, E12 AND E15

	Combined Samples		Beneficiaries		Participants	
	Percent Asked Question	Adjustment	Percent Asked Question	Adjustment	Percent Asked Question	Adjustment
Eligible for Question E3 (SSI Beneficiaries)						
All SSI Beneficiaries	60.8		61.5		63.9	
SSI Only	47.3	2.116	47.3	2.114	50.0	1.999
Both SSI and SSDI	86.0	1.163	87.9	1.138	81.2	1.232
Eligible for Question E12 (SSI Beneficiary Younger than 25 at Sample Selection and who Received Benefits before age 22)						
All SSI Eligible Beneficiaries	47.5		47.3		41.2	
SSI Only	53.1	1.884	52.8	1.895	56.0	1.784
Both SSI and SSDI	23.9	4.184	23.8	4.206	9.2	10.820
Eligible for Question E15 (SSDI beneficiaries)						
All SSDI Beneficiaries	45.0		45.0		40.4	
SSDI Only	55.5	1.802	55.5	1.802	52.1	1.918
Both SSI and SSDI	14.0	7.123	14.0	7.147	17.8	5.603

Source: NBS, round 1.

For question E3, the weighted percentage of people for which the question was asked was around 60 percent with 47 to 50 percent of the SSI only beneficiaries asked the question and 80 to 88 percent of the beneficiaries of both SSI and SSDI. For question E12, the eligible population were SSI beneficiaries who were younger than 25 at the time of sample selection and who also received benefits before age 22. For this question, the weighted percentages were lower overall (in the 40 to 48 percent range) and in the low to mid 50 percent range for SSI beneficiaries who were not also SSDI beneficiaries (adjustment factors less than 2.0), but less than 25 percent for the beneficiaries of both SSI and SSDI. The adjustment ratios for the weights of the beneficiaries of both SSI and SSDI were therefore in the range of 4 to 10. For question E15 (eligible population consisted of all SSDI beneficiaries), a similar pattern occurred and the overall

weighted percentages were lower with response rate in the low to mid 50 percent range for SSDI beneficiaries who were not also SSI beneficiaries, but less than 20 percent for the beneficiaries of both SSI and SSDI. The adjustment ratios for the weights of the beneficiaries of both SSI and SSDI were therefore in the range of 5 to 7. Because of the low response rates in some of the beneficiary groups, the data user is advised to be cautious with the use of estimates based on these questions.

IV. IMPUTATIONS

In the NBS, the data collection instruments were administered using computer-assisted interviewing (CAI) technology. The CAI technology allows the use of automated routing to move the respondent to the applicable questions and also implements checks of the entered data for consistency and reasonableness. In addition, because the program will not allow a question to be left blank, the interviewer cannot proceed unless an appropriate response has been entered ("don't know" and "refused" are included as response options and used as necessary). These processes substantially reduce the extent of item nonresponse for a complex survey, but some item nonresponse will still exist. Item nonresponse includes cases where the question was mistakenly not asked and cases where "don't know" or "refused" were recorded as responses.

For the NBS, imputation was used to compensate for item nonresponse. Two imputation methods were primarily used: deductive (or logical) imputation and unweighted hot-deck imputation. However, for some variables, insufficient data were available to use one of these two methods, so other specialized imputation procedures were employed to use the data available. The methods were selected based on the type of variable (dichotomous, categorical or continuous), the amount of missing data, and the availability of data for the imputations. For some variables, imputations were processed using a combination of methods.

Where possible, imputed values were made consistent with pre-existing nonmissing variables by excluding donors with potentially inconsistent imputed values. After each imputation was processed, the imputed values were evaluated using a variety of quality control procedures. If the initial imputed value was out of an acceptable range or inconsistent with other data for that case, the imputation was repeated until the imputed value was in range and consistent with other reported data.

Deductive, or logical, imputation is the assignment of a value that can be deduced from other data, or for which there is a high degree of certainty that the value is correct. This method was based on a review of data related to the imputed variable.

The hot-deck imputation procedure entails the classification of sample members into mutually exclusive and exhaustive imputation classes (or imputation cells) of respondents who are assumed to be similar relative to the key population variables (such as age, disability status, and SSI recipient status). For each sample member with a missing value (a recipient), a sample member with complete data (a donor) is chosen within the same imputation class to provide a value. It is desirable to have the imputation class contain sufficient sample members to avoid the selection of a single donor for multiple sample members with missing data. The hot-deck procedure is computationally efficient and, in a recent National Center for Education Statistics working paper (USDE 2001), a simulation study showed that a hot-deck procedure fared well in comparison to more sophisticated imputation procedures, including multiple imputation, Bayesian bootstrap imputation, and ratio imputation. However, it should be noted that no attempt was made to estimate the component of variance due to imputation, even though such a component is always positive. Users should be aware that variance estimates using imputed data will be underestimates, with the amount of bias in the variance estimate directly related to the amount of missingness in the variable of interest. For most of the variables requiring imputation, the extent of missingness was low, so that this component would be very small.

The hot-deck imputation procedure used an unweighted selection process to select a donor, with selections done within imputation classes defined by key related variables for each application. This was accomplished in two ways. In one of the applications, in addition to the variables defining the imputation classes, a sorting variable was included where the recipient and all donors within the imputation class were sorted together by the levels of this variable. Using

the sorted data within the imputation class, a case immediately preceding or following a sample member with missing data was randomly selected as the donor with equal probability. In the other application, a donor was randomly selected from within the imputation class. With either method, we allowed with-replacement selection of a donor for each recipient. In other words, a sample member could have been a donor for more than one recipient. Because the extent of missing values was very low, only a few donors were used more than once.

The factors used to form the cells for each imputed variable needed to be appropriate for the population, the data collected, and the purpose of the study. The imputation classes also needed to have a sufficient count of donors for each sample member with missing data. We used a variety of methods to form the imputation classes. These methods included bivariate cross-tabulations, step-wise regressions, and multivariate procedures such as CHAID (Chi-squared Automatic Interaction Detection software attributed to Kass [1980] and Biggs, et al. [1991], and its application in SPSS is described in Magidson [1993]). To develop these imputation classes, we used information from both the interview and SSA data files. Classing and sorting variables were closely related to the variable being imputed (the response variable). Sorting variables were either less closely related to the response variable than classing variables, or were forms of the classing variables with finer levels. As an example of the latter situation, four age categories were sometimes used as imputation classes: (1) 18 to 29, (2) 30 to 39, (3) 40 to 49, and (4) 50 to 64. The actual age could then be used as a sorting variable, so that donors and recipients were as close together as possible in age.

The hot-deck software automatically identified situations where the imputation class only contained recipients and no donors. In these cases, imputation classes were collapsed and the imputation redone using the collapsed classes. The strategy for collapsing classes required a ranking of the variables used to define the imputation class with regard to each variable's

relationship to the variable requiring imputation. Those variables less closely related to the variable requiring imputation were more likely to have levels collapsed. In addition, variables with many levels were also more likely to have levels collapsed. In general, if more than a very small number of imputation classes required collapsing, then one or more variables were dropped from the definition of the imputation class and the imputation procedure was rerun.

Some variables were constructed from two or more variables. For some of the "constructed" variables, it was more efficient to impute the component variables and then to impose the recoding of the constructed variable on these imputed values. These component variables are not shown in the following tables because they were not included in the final data set.

For some of the imputed variables in the data set, the number of missing responses does not match the number of imputed responses. Often, these variables correspond to questions that follow a filter question. For example, question I33 asks if the respondent has difficulty climbing 10 steps and the follow-up question if the response is "yes", I34, asks if the respondent is able to climb 10 steps at all. In order to be asked the follow-up question, the respondent must have answered "yes" to the screener question. If the respondent answered "no," the follow-up question was coded a legitimate missing (".l") which was not imputed. However, if the respondent refused to answer the screener question, the follow-up question was also coded a legitimate missing. If the screener variable was then imputed to be "yes," the response to the follow-up question was imputed. This caused the count of the actual number of imputed responses to be greater than the number of missing or invalid responses.

A. NBS IMPUTATIONS OF SPECIFIC VARIABLES

Included below in several tables is information about how imputation was employed in the NBS. The tables include the imputed variable names and a brief description of each. The tables also include the methods of imputation, total number of missing responses, the number of

respondents eligible for the question, and the percentage of responses imputed. This information is recorded on the final file with an imputation flag, identified by the suffix "iflag", which has the following eight levels: (.) legitimate missing or no answer, (0) self-reported data, (1) logical imputation, (2) administrative data, (3) hot-deck imputed, (4) imputation using the distribution of a variable related to the variable being imputed, (5) imputation based on specialized procedures specific to Section K, and (6) constructed from other variables with imputed values. In most cases, the logical assignments were done using imputed values.[25] Therefore, the distinction between "logically assigned" and "constructed from other variables with imputed values" is somewhat opaque. In general, if a logical assignment is done for variables corresponding directly to questionnaire questions, the flag is set to 1. For variables constructed from these variables (prefixed with a "C_"), the flag is set to 6.

In the sections that follow, summaries of the imputations conducted are given, organized by the sections within the questionnaire to which the variables correspond. Details of some of the imputation types are given for each section.

1. Section L: Race and Ethnicity

Several questions included on the NBS instrument gathered information on the respondents' race and ethnicity. Two of these variables, located in Section L, include imputed responses and are described in Table IV.1. In particular, L1_i corresponds to the question asking whether the respondent is Hispanic or not; C_Race_i corresponds to the question asking about the respondent's race.

[25] No distinction was made between logical assignments using imputed values and logical assignments using self-reported values.

TABLE IV.1

RACE AND ETHNICITY IMPUTATIONS

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
L1_i	Hispanic/Latino Ethnic Origins	1 logical imputation, 10 imputations from SSA's administrative data, 124 imputations from random hot deck	135	7,603	1.78
C_Race_i	Race	116 imputations from SSA's administrative data, 176 imputations from random hot deck	292	7,603	3.84

Source: NBS, round 1.

In this table, respondents who did not indicate in the questionnaire whether they were Hispanic were classified as such if the SSA administrative data so indicated; the single logical imputation was conducted by looking at the name of the respondent and comparing it to a list of Hispanic names provided by the North American Association of Central Cancer Registries (NAACCR 2003). For respondents who still had missing data, the Hispanic indicator was imputed using a random hot deck using imputation classes defined by the zip code of each sample member.

Respondents could choose from five race categories: white, black/African American, Asian, Hawaiian/Pacific Islander, and Native American/American Indian. Respondents were allowed to select more than one of these categories to identify themselves (as prescribed by the Office of Management and Budget). The final race variable on which imputation was applied had six categories. Respondents who indicated Hawaiian/Pacific Islander were included with respondents reporting multiple races other than black/African American. A separate category was used for respondents reporting multiple races including black/African American. Although the SSA administrative data did not have a category for multiple races, respondents with race information in the SSA files were categorized according to four of the five categories above

(Hawaiian/Pacific Islanders were included with the respondents reporting Asian). Respondents who did not answer the race question but did have race information in the SSA files were categorized into one of the four categories. This resulted in misclassification of respondents with extant SSA administrative data who didn't answer the race question in the survey, but would have identified themselves in the survey as multiple race or Hawaiian/Pacific Islander. Hawaiian/Pacific Islanders were presumably misclassified as Asian using SSA administrative data. However, we assumed that the number of respondents like this was small so that misclassification was not a major problem. As with the Hispanic indicator, for respondents that still had missing data, race was imputed using a random hot deck using imputation classes defined by the zip code of each sample member.

2. Section B: Disability Status Variables and Work Indicator

Table IV.2 describes five imputed variables that pertain to the sample member's disability status and an indicator whether the respondent was currently working. These imputed variables include three variables which collapse and recode primary diagnosis codes from the International Classification of Diseases, Ninth Revision (ICD-9) in three different ways (C_MainConBodyGroup_i, which corresponds to the collapsing done in Table II.2, C_MainConDiagGrp_i, and C_MainConColDiagGrp_i), age when the disability was first diagnosed (C_DisAge_i), and an indicator of childhood or adult onset of the disability (C_AdultChildOnset_i). A fourth variable with collapsed primary diagnosis codes was also imputed, with levels further collapsed from C_MainConDiagGrp_i. This variable (C_MainConImput_i) is not included in Table IV.2 because it was not released to the final file, but it was used in subsequent imputations as a classing variable. All missing values for C_AdultChildOnset_i were "logically assigned" using the imputed values from C_DisAge_i, the age-of-onset variable. In addition, Section B contains a question asking whether the respondent

was currently working (B24_i). This question is a gate question for all of the work status variables in Section C.

TABLE IV.2

DISABILITY STATUS IMPUTATIONS

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
C_MainConDiagGrp_i	Primary diagnosis group	75 sequential hot deck	75[a]	7,043	1.06
C_MainConColDiagGrp_i	Main condition diagnosis group collapsed	75 sequential hot deck	75[a]	7,043	1.06
C_MainConBodyGroup_i	Main condition body group	75 sequential hot deck	75[a]	7,043	1.06
C_Disage_i	Age at onset of disability	240 sequential hot deck	240	7,603	3.16
C_Adultchild_onset_i	Adult/child onset of disability	23 logical	23	7,603	0.30
B24_i	Currently working	10 random hot deck	10	7,603	0.13

Source: NBS, round 1.

[a] Includes 20 cases coded as don't know or refused on B1 and cases coded as don't know, refused, no condition reported, and uncodeable for diagnosis code variables.

For variables where hot-deck imputation was required, the sequential hot deck with a sorting variable was used for the recoded and collapsed diagnosis codes, as well as disability age. The work indicator variable used a random hot deck. All of the variables in Section B used an indicator of whether the onset of the disability was in childhood or adulthood, as well as age and gender to define imputation classes. One of the collapsed condition code variables, C_MainConImput_i was also used as a classing variable for disability age and the work indicator. Additional classing variables were used that were specific to the variable being imputed.

3. Section C: Current Jobs Variables

Several questions in the National Beneficiary Survey asked respondents about current employment. In Section C, these questions were only asked of respondents who indicated that they were currently working in question B24. They include salary (C_MainCurJobHrPay_i, C_MainCurJobMnthPay_i, and C_TotCurJobMnthPay_i), usual hours worked at the job or jobs (C8_1_i, C_TotCurWkHrs_i, and C_TotCurHrMnth_i), the number of places the respondent was employed (C1_i), and job description of the place of main employment (C2_1_1d_i). These variables are identified in Table IV.3.

TABLE IV.3

CURRENT JOBS IMPUTATIONS

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
C1_i	Number of current jobs	1 random hot deck	1	1,155	0.09
C2_1_1d_i	Main current job SOC code to one digit	12 random hot deck	12[a]	1,155	1.04
C8_1_i	Hours per week usually worked at current main job	35 random hot deck	35	1,155	3.03
C_TotCurWkHrs_i	Total weekly hours at all current jobs	35 logical, 4 constructed from imputed variables	39[b]	1,155	3.38
C_TotCurHrMnth_i	Total hours per month at all current jobs	39 constructed from imputed variables	39	1,155	3.38
C_MainCurJobHrPay_i	Hourly pay at current main job	145 constructed from imputed variables	145	1,155	12.55
C_MainCurJobMnthPay_i	Monthly pay at current main job	121 logical, 29 imputed by distributional assumptions, 5 constructed from imputed variables	155	1,155	13.42
C_TotCurMnthPay_i	Total monthly salary all current jobs	29 logical, 126 sequential hot deck, 7 constructed from imputed variables	162	1,155	14.03

Source: NBS, round 1.

[a] Includes 10 cases coded as don't know or refused on B24 which were imputed at B24_i.

[b] The 39 missing values do not include one case where the number of jobs was imputed to 1, but the number of hours at the main job was not missing. The flag for the total number of hours worked in this case was set to 0 ("self-reported"). The same is true for the missing values in the other total composite variables (C_TotCurHrMnth.and C_TotCurMnthPay)

Some of the variables in this table had missing values that were not directly imputed. Rather, constituent variables not included in this table had missing values that were imputed, and then combined to form the variables below. For example, C_TotCurWkHrs_i was constructed from the number of hours per week usually worked at the current main job plus the number of hours for each of the respondent's other jobs. In most cases, the respondent worked one job so C_TotCurWkHrs_i was set equal to C8_1_i. However, if the respondent worked multiple jobs, and the number of hours in secondary jobs was imputed, then C_TotCurWkHrs_i was "constructed from imputed variables."

Other variables had values imputed by using the distribution of a variable related to the variable at hand. For example, if the take-home monthly pay of the respondent's current main job was not missing but the gross monthly pay (C_MainCurJobMnthPay_i) of the respondent's current main job was missing, then the relationship between gross monthly pay and take-home monthly pay among respondents missing neither variable was used to determine the appropriate value for gross monthly pay. In particular, a random draw was selected from the observed distribution of relative taxes, where "relative tax" is defined as the proportion of a respondent's pay devoted to tax. This randomly drawn relative tax was used to determine an imputed gross monthly pay for 29 cases with missing data for C_MainCurJobMnthPay. As Table IV.3 indicates, hot-deck imputations were only applied to four of the jobs variables: C1_i, C2_1_1d_i, C8_1_i, and C_TotCurMnthPay_i. For C1_i, C2_1_1d_i, and C8_1_i, a random hot deck was used, with the collapsed condition code variable and level of education used as classing variables. Additional classing variables were also used that were specific to each variable. The sequential hot deck with a sorting variable was used in the imputation of missing values for C_TotCurMnthPay_i. The classing variables for this imputation were education, total number of

hours worked on current jobs, collapsed job description code, and number of jobs, with the collapsed condition code variable used as a sorting variable.

4. Section I: Health Status Variables

A total of 56 health status variables are in Section I of the National Beneficiary Survey questionnaire where imputations were applied. The 56 imputed variables in this section, and the methods of imputation used in each case, are identified in Table IV.4. These items cover a range of topics, from the respondent's general health to more specific questions on the instrumental activities of daily living (IADLs) and activities of daily living (ADLs) and other health and coping indicators. Also included in this section are a series of questions pertaining to the respondent's use of illicit drugs and alcohol.

TABLE IV.4

HEALTH STATUS IMPUTATIONS

Variable Name	Description	Imputation Method[a]	Number Missing	Number Eligible	Percent Imputed
I1_i	Health during the past four weeks	21 hot deck	21	7,603	0.28
I9_i	Current health	51 hot deck	51	7,603	0.67
I17_i	Difficulty seeing	4 logical, 83 hot deck	87	7,603	1.14
I18_i	Able to see at all	50 logical, 59 hot deck	109	3,225	3.38
I19_i	Uses special equipment because of difficulty seeing	50 logical, 12 hot deck	62	3,225	1.92
I21_i	Difficulty hearing	1 logical, 35 hot deck	36	7,603	0.47
I22_i	Able to hear normal conversation	27 logical, 13 hot deck	40	1,393	2.87
I23_i	Uses special equipment because of difficulty hearing	27 logical, 3 hot deck	30	1,393	2.15

TABLE IV.4 *(continued)*

Variable Name	Description	Imputation Method[a]	Number Missing	Number Eligible	Percent Imputed
I25_i	Difficulty having speech understood	6 logical, 40 hot deck	46	7,603	0.61
I26_i	Able to have speech understood at all	26 logical, 14 hot deck	40	2,285	1.75
I27_i	Uses special equipment because of difficulty speaking	26 logical, 3 hot deck	29	2,285	1.27
I29_i	Difficulty walking without assistance	16 logical, 57 hot deck	73	7,603	1.03
I30_i	Able to walk ¼ mile	40 logical, 64 hot deck	104	3,561	2.92
I31_i	Uses special equipment because of difficulty walking	40 logical, 10 hot deck	50	3,561	1.40
I33_i	Difficulty climbing 10 steps	10 logical, 93 hot deck	103	7,603	1.35
I34_i	Able to climb 10 steps at all	52 logical, 60 hot deck	112	3,682	3.04
I35_i	Difficulty lifting and carrying 10 lbs.	3 logical, 61 hot deck	64	7,603	0.84
I36_i	Able to lift or carry 10 lbs. at all	35 logical, 70 hot deck	105	3,443	3.05
I37_i	Difficulty using hands or fingers	32 hot deck	32	7,603	0.42
I38_i	Able to use hands or fingers at all	27 logical, 20 hot deck	47	1,994	2.63
I39_i	Difficulty reaching over head	1 logical, 49 hot deck	50	7,603	0.66
I40_i	Able to reach over head at all	36 logical, 22 hot deck	58	2,049	2.83
I41_i	Difficulty standing	1 logical, 72 hot deck	73	7,603	0.96
I42_i	Able to stand at all	32 logical, 17 hot deck	49	4,519	1.08
I43_i	Difficulty stooping	3 logical, 56 hot deck	59	7,603	0.78
I44_i	Able to stoop at all	34 logical, 38 hot deck	72	4,335	1.66
I45_i	Difficulty getting around inside home	7 logical, 28 hot deck	35	7,603	0.46
I46_i	Need help to get around inside home	24 logical, 16 hot deck	40	1,320	3.03

TABLE IV.4 *(continued)*

Variable Name	Description	Imputation Method[a]	Number Missing	Number Eligible	Percent Imputed
I47_i	Difficulty getting around inside home	9 logical, 42 hot deck	51	7,603	0.67
I48_i	Need help to get around outside home	23 logical, 33 hot deck	56	3,057	1.83
I49_i	Difficulty getting into/out of bed	2 logical, 38 hot deck	40	7,603	0.53
I50_i	Need help getting into/out of bed	26 logical, 32 hot deck	58	2,140	2.71
I51_i	Difficulty bathing or dressing	5 logical, 34 hot deck	39	7,603	0.51
I52_i	Need help bathing or dressing	27 logical, 15 hot deck	42	1,862	2.26
I53_i	Difficulty shopping	30 logical, 32 hot deck	62	7,603	0.82
I54_i	Need help shopping	25 logical, 15 hot deck	40	2,664	1.50
I55_i	Difficulty preparing own meals	18 logical, 38 hot deck	56	7,603	0.74
I56_i	Need help to prepare meals	25 logical, 17 hot deck	42	2,895	1.45
I57_i	Difficulty eating	29 hot deck	29	7,603	0.38
I58_i	Need help to eat	26 logical, 6 hot deck	32	1,067	3.00
I59_i	Trouble concentrating	86 hot deck	86	7,603	1.13
I60_i	Trouble coping with stress	117 hot deck	117	7,603	1.54
I61_i	Trouble getting along with people	96 hot deck	96	7,603	1.26
C_EquipFuncLim_i	Use equipment/device for functional/sensory limitation	16 constructed from imputed variables	16	7,603	0.21
C_NumSenLim_i	Number of sensory limitations	134 constructed from imputed variables	134	7,603	1.76
C_NumSevSenLim_i	Number of severe sensory limitations	101 constructed from imputed variables	101	7,603	1.33
C_NumPhyLim_i	Number of physical functional limitations	280 constructed from imputed variables	280	7,603	3.68

TABLE IV.4 *(continued)*

Variable Name	Description	Imputation Method[a]	Number Missing	Number Eligible	Percent Imputed
C_NumSevPhyLim_i	Number of severe physical functional limitations	270 constructed from imputed variables	270	7,603	3.55
C_NumEmotLim_i	Number of emotional/social limitations	226 constructed from imputed variables	226	7,603	2.97
C_NumADLs_i	Number of impaired activities of daily living (ADLs)	70 constructed from imputed variables	70	7,603	0.92
C_NumADLAssist_i	Number of ADLs requiring assistance	69 constructed from imputed variables	69	7,603	0.91
C_NumIADLs_i	Number of instrumental activities of daily living (IADL) difficulties	117 constructed from imputed variables	117	7,603	1.50
C_NumIADLAssist_i	Number of IADLs Requiring Assistance	68 constructed from imputed variables	68	7,603	0.89
C_PCS8TOT_i	Physical summary score	321 constructed from imputed variables	321	7,603	4.22
C_MCS8TOT_i	Mental summary score	321 constructed from imputed variables	321	7,603	4.22
CageScore_indicator_i	CAGE Alcohol Score	39 constructed from imputed variables	39	7,603	0.51
I72_i	Use drugs in larger amounts than prescribed	49 hot deck	49	7,603	0.64
C_DrugDep_i	Drug dependence	56 constructed from imputed variables	56	7,603	0.74

Source: NBS, round 1.

[a]For all of the imputations using hot deck in this section, a random hot deck was used. There was therefore no need to distinguish between random and sequential hot decks.

An example of a logical assignment in this section: if a respondent did not answer whether they had difficulty seeing newsprint letters (I17), but indicated that they couldn't see newsprint letters at all (I18) or required special devices to read newsprint letters (I19), then I17_i was logically assigned "yes."

As in previous sections, "constructed from imputed variables" refers to the fact that the constituent variables of each constructed variable were imputed.

All of the variables requiring imputation of missing values in the Health Status section were imputed using a random hot deck. The only classing variable that was common to all imputations was the collapsed condition code variable. Age and gender were also used in most imputations. The remainder of classing and sorting variables was specific to the variable being imputed.

5. Section K: Sources of Income Other than Employment

The imputed variables presented in this section are constructed variables that pertain to nonemployment-based income. These include worker's compensation, private disability claims, unemployment, and generally other sources of regular income. The imputed variables in this section are described in Table IV.5.

TABLE IV.5

IMPUTATIONS ON SOURCES OF INCOME OTHER THAN EMPLOYMENT

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
C_AmtPrivDis_i	Amount received from private disability last month	85 logical, 37 imputed using specialized procedures	122	7,603	1.60
C_AmtWorkComp_i	Amount received from workers' compensation last month	40 logical, 3 imputed using specialized procedures	43	7,603	0.57
C_AmtVetBen_i	Amount received from veterans' benefits last month	35 logical, 18 imputed using specialized procedures	53	7,603	0.68
C_AmtPubAssis_i	Amount received from public assistance last month	60 logical, 35 imputed using specialized procedures	95	7,603	1.25
C_AmtUnemply_i	Amount received from unemployment benefits last month	38 logical, 1 imputed using specialized procedures	39	7,603	0.51

TABLE IV.5 *(continued)*

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
C_AmtPrivPen_i	Amount received from private pension last month	42 logical, 16 imputed using specialized procedures	58	7,603	0.76
C_AmtOthReg_i	Amount received from other regular sources last month	49 logical, 24 imputed using specialized procedures	73	7,603	0.96

Source: NBS, round 1.

In this section, respondents were first asked if they had received money from a specific source and then for the specific amount received from that source. If a respondent could not provide a specific value, he or she was asked a series of questions on whether the value was above or below specific values, or was given the option of providing a range of values, where the optional ranges depended upon responses to a series of questions. After being classified according to a range of values that he or she provided, the respondent was assigned the median of the specific values provided by respondents who gave responses within the same range. If a respondent could not say whether the actual value was above or below a specific threshold, we imputed first the range (using a random assignment) and then assigned the median of the values provided by respondents who gave specific values within that range. If the respondent did not know if they received funds from a source, we then imputed whether they did or did not using a random hot deck, and then proceeded as above.

The logical assignments in this section derive from imputed values in the constituent questions. For example, if the respondent was imputed to not have received private disability insurance (K6a_i), then C_AmtPrivDis_i was a logically assigned "no." Otherwise, if any income was derived from these sources but an imputation was required at some point in the sequence (either everything was imputed, or just the individual income was imputed) then the imputation flag indicated imputation by "special procedures."

For variables requiring hot-deck imputation, a random hot deck was used for all imputations. The classing variables were the same for all variables: an indicator of whether the respondent was a recipient of SSI, SSDI, or both; living situation; and education. None of the variables requiring hot-deck imputation are listed in Table IV.5 because they were only component variables for the delivered variables listed in the table.

6. Section L: Personal and Household Characteristics

Other than the personal characteristics of race and ethnicity discussed earlier, most of the imputed variables in section L pertain to household characteristics. These questions include education (L3_i), marital status (L8_i), cohabitation status (C_Cohab_i), number of children in the household (C_NumChildHH_i), household size (C_Hhsize_i), and poverty level (FedPovertyLevel_cat1). Also included in this section is the constructed variable for the respondent's body mass index (C_BMI_cat_i), since it is constructed of variables collected in section L. Most of these variables were imputed early in imputation processing and were used in the imputation of work status variables; however, poverty level was imputed later. Both sets of variables are discussed in this section.

The imputation of poverty level required the imputation of annual income and household size. The annual income question was another case of a specific value being requested, and if a specific value could not be provided, the respondent was asked if the annual income fell in certain ranges. For this item, some respondents provided a specific value; some respondents answered the questions on the ranges, and some refused to provide any information. Although annual income was a key variable used in the imputation of poverty level, it is not included in this table since it was not released in the final file. All of the missing values in

C_FedPovertyLevel_cat1[26] were derived from the imputed annual incomes; hence all missing values are "constructed from imputed variables." Table IV.6 identifies imputed variables in section L.

TABLE IV.6

IMPUTATIONS OF PERSONAL AND HOUSEHOLD CHARACTERISTICS

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
C_BMI_Cat_i	Body Mass Index categories	6 logical, 286 hot deck	292	7,603	3.84
L3_i	Highest year/grade completed in school	162 hot deck	162	7,603	2.13
L8_i	Marital Status	53 hot deck	53	7,603	0.70
L11_i	Living arrangements	27 logical, 14 hot deck	41	7,603	0.54
C_NumChildhh_i	Number of children living in the household	6 logical, 12 hot deck	18	7,603	0.24
C_hhsize_i	Household Size	24 logical, 20 hot deck	44	7,603	0.58
C_cohab_i	Cohabitation Status	6 logical, 43 hot deck	49	7,603	0.64
C_FedPovertyLevel_cat1	2003 Federal Poverty Level	2,754 constructed from imputed variables	2,754	7,603	36.22

Source: NBS, round 1.

Logical assignments in this section are based on related variables also in this section. For example, the 27 logical assignments for L11_i are due to the fact that 27 respondents did not answer L11, but indicated in L16 that only one adult lived in the household.

For all of the variables requiring hot-deck imputation that are listed in Table IV.6, a random hot deck was used. The only classing variable common to all imputations was the collapsed

[26] The name of this variable reflects that fact that the final variable was a categorical (as opposed to a continuous) measure of poverty levels.

condition code variable. Other variables were specific to the variable being imputed. The imputed annual incomes that were used in the determination of C_FedPovertyLevel_cat1 were imputed using a sequential hot deck with a sorting variable.

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

V. ESTIMATING SAMPLING VARIANCE FOR NBS

The sampling variance of an estimate derived from survey data for a statistic (such as a total, a mean or proportion, or a regression coefficient) is a measure of the random variation among estimates of the same statistic computed over repeated implementation of the same sample design with the same sample size on the same population. The sampling variance is a function of the population characteristics, the form of the statistic, and the nature of the sampling design. The two general forms of statistics are linear combinations of the survey data (for example, a total) and nonlinear combinations of the survey data. Nonlinear combinations include the ratio of two estimates (for example, a mean or a proportion in which both the numerator and the denominator are estimated) and more complex combinations such as regression coefficients. For linear estimates with simple sample designs (such as a stratified or unstratified simple random sample) or complex designs (such as stratified multistage designs), explicit equations are available to compute the sampling variance. For the more common nonlinear estimates with simple or complex sample designs, explicit equations are not generally available and various approximations or computational algorithms are used to provide an essentially unbiased estimate of the sampling variance.

The NBS sample design involves stratification and unequal probabilities of selection. Variance estimates calculated from NBS data must incorporate the sample design features in order to obtain the correct estimate. Most procedures in standard statistical packages, such as SAS and SPSS, are not appropriate for analyzing data from complex survey designs, such as the NBS design. These procedures assume independent, identically distributed observations or simple random sampling with replacement. Although the simple random sample (SRS) variance may approximate the true sampling variance for some surveys, it is likely to substantially

underestimate the sampling variance with a design as complex as the NBS design. Complex sample designs have led to the development of a variety of software options that require the user to identify essential design variables such as strata, clusters, and weights. [27]

The most appropriate sampling variance estimators for complex sample designs such as the NBS are the procedures based on the Taylor series linearization of the nonlinear estimator using explicit sampling variance equations and the procedures based on forming pseudo-replications[28] of the sample. The Taylor series linearization procedure is based on a classic statistical method in which a nonlinear statistic can be approximated by a linear combination of the components within the statistic. The accuracy of the approximation is dependent on the sample size and the complexity of the statistic. For most commonly used nonlinear statistics (such as ratios, means, proportions, and regression coefficients), the linearized form has been developed and has good statistical properties. Once a linearized form of an estimate is developed, the explicit equations for linear estimates can be used to estimate the sampling variance. Because the explicit equations can be used, the sampling variance can be estimated using many of the features of the sampling design (for example, finite population corrections, stratification, multiple stages of selection, and unequal selection rates within strata). This is the basic variance estimation procedure used in SUDAAN, the survey procedures in SAS, Stata, and other software packages to accommodate simple and complex sampling designs. To be able to calculate the variance, sample design information (such as stratum, analysis weight, and so on) is needed for each sample unit.

[27] A website that reviews software for variance estimation from complex surveys, created with the encouragement of the Section on Survey Research Methods of the American Statistical Association, is available on-line at http://www.fas.harvard.edu/~stats/survey-soft/survey-soft.html. The site lists software packages available for personal computers and provides direct links to the home pages of these packages. The site also contains articles and links to articles that provide general information about variance estimation and links to articles that compare features of the software packages.

[28] Pseudo-replications are restricted or random subsamples of a specific survey sample, as opposed to true replications of the sampling design, which entails the selection of multiple independent samples using the same sampling design.

Currently, more survey data analysis software packages use the Taylor series linearization procedure and explicit sampling variance equations. Therefore, we developed the variance estimation specifications necessary for the Taylor series linearization procedure (PseudoStrata and PseudoPSU). Example code for this procedure using SAS and the survey data analysis software SUDAAN is given in Appendix E.[29] Details about syntax for SAS are available from SAS (SAS Institute 2004). Details about SUDAAN syntax are available from RTI International (Research Triangle Institute 2004).

[29] The example code provided in Appendix E is for simple descriptive statistics using the procedures DESCRIPT in SUDAAN and SURVEYMEANS in SAS. Other procedures in SAS (SURVEYREG, SURVEYFREQ, and SURVEYLOGISTIC) and in SUDAAN (CROSSTAB, REGRESS, LOGISTIC, MULTILOG, LOGLINK, and SURVIVAL) are available for more complex analyses. Since SUDAAN was created specifically for survey data, the range of analyses that can be performed with these data in SUDAAN is much wider than in SAS.

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

VI. REFERENCES

Bethel, J. and D. Stapleton. "Evaluation Design for the Ticket to Work Program: Final Survey Sample Design." Report submitted to the Social Security Administration. Falls Church, VA: The Lewin Group, July 2002.

Biggs, D., B. deVille, and E. Suen. "A Method of Choosing Multiway Partitions for Classification and Decision Trees," Journal of Applied Statistics, vol. 18, 1991, pp. 49-62.

Cox, D. R. and E. J. Snell. *The Analysis of Binary Data*, Second Edition. London: Chapman and Hall, 1989.

Folsom, R.E., F.J. Potter, and S.R. Williams. "Notes on a Composite Size Measure for Self to Weighting Samples in Multiple Domains." *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 1987, pp. 792-796.

Kass, G. V. "An Exploratory Technique for Investigating Large Quantities of Categorical Data," Applied Statistics, vol. 29, 1980, pp. 119-127.

Magidson, J. "SPSS for Windows CHAID Release 6.0." Belmont MA: Statistical Innovations, Inc., 1993.

NAACCR Expert Panel on Hispanic Identification. "Report of the NAACCR Expert Panel on Hispanic Identification 2003." Springfield, IL: North American Association of Central Cancer Registries. 2003.

Research Triangle Institute. *SUDAAN Language Manual, Release 9.0*. Research Triangle Park, NC: Research Triangle Institute. 2004.

SAS® Institute. *SAS/STAT® 9.1 User's Guide*. Cary, NC: SAS Institute. 2004.

Thornton, C., Livermore, G., Stapleton, D., Kregel, J., Silva, T., O'Day, B., Fraker, T., Revell, Jr., W., Schroeder, H., and Edwards, M. "Evaluation of the Ticket to Work Program Initial Evaluation Report." Washington, DC: Mathematica Policy Research, Inc. January 2004.

U.S. Department of Education. National Center for Education Statistics. "A Study of Imputation Algorithms." Working Paper No. 2001-17. Ming-xiu and Sameena Salvucci. Washington, DC. 2001.

Wright, D. and K. Barrett. "National Beneficiary Survey Round 1 (Volume 2 of 3): Data Cleaning and Identification of Data Problems". Report submitted to the Social Security Administration. Washington, DC: Mathematica Policy Research, Inc., December 2008.

Wright, D., M. Sloan, F. Potter, E. Grau, K. Barrett, and M. Cidade. "National Beneficiary Survey: Round 1 (Volume 3 of 3): User's Guide for Restricted and Public Use Files." Report submitted to the Social Security Administration. Washington, DC: Mathematica Policy Research, Inc., December 2008.

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

APPENDIX A

OTHER/SPECIFY AND OPEN-ENDED ITEMS WITH ADDITIONAL CATEGORIES CREATED DURING CODING

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

OTHER/SPECIFY AND OPEN-ENDED ITEMS WITH ADDITIONAL CATEGORIES CREATED DURING CODING

Item	Question Text	Questionnaire Response Options	Categories Added During Coding
B25	What are they (the other reasons you are not working that I didn't mention)?	a=A physical or mental condition prevents {you/him/her} from working b={You/name} cannot find a job that {you are/(he/she) is} qualified for c={You do/name does} not have reliable transportation to and from work d={You are/name is} caring for someone else f={You/name} cannot find a job {you want/(he/she) wants} g={You are/name is} waiting to finish school or a training program h=Workplaces are not accessible to people with {your/name's} disability i={You do/name does} not want to lose benefits such as Disability, Worker's Compensation, or Medicaid j={Your/name's} previous attempts to work have been discouraging l=Others do not think {you/name} can work m=Employers will not give {you/name} a chance to show that {you/he/she} can work	n=Can't find a job/job market is bad o=Lack skills
B39	Who {do you/does NAME} discuss your work goals with the most?	01=PARENT/GUARDIAN 02=SPOUSE/PARTNER 03=FRIEND 04=JOB COACH 05=EMPLOYER/SUPERVISOR 06=OTHER RELATIVE 07=CASE WORKER/COUNSELOR/PROGRAM STAFF 08=MEDICAL PROVIDER 09=OTHER (SPECIFY: <OPEN>)	10=OTHER NON-RELATIVE

Item	Question Text	Questionnaire Response Options	Categories Added During Coding
B42	Who else {do you/does NAME} discuss {your/his/her} work goals with?	01=PARENT/GUARDIAN 02=SPOUSE/PARTNER 03=FRIEND 04=JOB COACH 05=EMPLOYER/SUPERVISOR 06=OTHER RELATIVE 07=CASE WORKER/COUNSELOR/PROGRAM STAFF 08=MEDICAL PROVIDER 09=OTHER (SPECIFY: <OPEN>)	10=OTHER NON-RELATIVE
B45	Who else {do you/does NAME} discuss {your/his/her} work goals with?	01=PARENT/GUARDIAN 02=SPOUSE/PARTNER 03=FRIEND 04=JOB COACH 05=EMPLOYER/SUPERVISOR 06=OTHER RELATIVE 07=CASE WORKER/COUNSELOR/PROGRAM STAFF 08=MEDICAL PROVIDER 09=OTHER (SPECIFY: <OPEN>)	10=OTHER NON-RELATIVE
C23	What kind of special equipment {do you/does NAME} use?	01=BRACE 02=CANE/CRUTCHES/WALKER 03=WHEELCHAIR 04=MODIFIED COMPUTER HARDWARE 05=MODIFIED COMPUTER SOFTWARE 06=OTHER (SPECIFY: <OPEN>)	07=HEARING AIDS 08=GLASSES
C35	Are there any changes in {your/NAME's} {main/current} job or workplace related to {your/his/her} mental or physical condition that {you need/he/she needs}, but that have not been made? (IF YES) What are those changes?	<OPEN>	a=Need special equipment b=Need changes in work schedule c=Need changes to the tasks d=Need changes to environment e=Need co-workers to assist

Item	Question Text	Questionnaire Response Options	Categories Added During Coding
D23	Why did {you/NAME} stop working at this job?	LAYOFF, FIRED, RETIRED 01=LAYOFF, PLANT CLOSED 02=FIRED 03=RETIRED/OLD AGE 04=JOB WAS TEMPORARY AND ENDED PROBLEMS WITH JOB 05=DID NOT LIKE SUPERVISOR OR CO-WORKERS 06=DID NOT LIKE JOB DUTIES 07=DID NOT LIKE JOB EARNINGS 08=DID NOT LIKE BENEFITS 09=DID NOT LIKE OPPORTUNITIES FOR ADVANCEMENT 10=DID NOT LIKE LOCATION 11=DID NOT GET ACCOMMODATIONS THAT WERE NEEDED OTHER PROBLEMS 12=TRANSPORTATION PROBLEMS 13=DECIDED TO GO TO SCHOOL 14=CHILD CARE RESPONSIBILITIES (PREGNANT) 15=OTHER FAMILY OR PERSONAL REASONS DISABILITY 16=DISABILITY GOT WORSE 17=BECAME DISABLED 18=OTHER (SPECIFY: <OPEN>)	19=MOVED TO ANOTHER AREA 20=FOUND ANOTHER JOB 21=LOSS OR POTENTIAL LOSS OF BENEFITS 22=WORK SCHEDULE
D25	Did you work fewer hours or earn less money than you could have because {you/he/she} you…	a={Were/Was} taking care of somebody else b={Were/Was} enrolled in school or a training program c=Wanted to keep Medicare or Medicaid coverage d=Wanted to keep cash benefits such as disability or workers compensation e=Just didn't want to work more f=Are there any reasons I didn't mention why {you/NAME} might have chosen to work or earn less than {you/he/she} could have during 2004 (SPECIFY: <OPEN>)	g=Had medical problems/complications

Item	Question Text	Questionnaire Response Options	Categories Added During Coding
D26	In 2004, do you think {you/NAME} could have worked or earned more if {you/he/she} had:	a=Help caring for {your/his/her} children or others in the household b=Help with {your/his/her} own personal care c=Reliable transportation to and from work d=Better job skills e=A job with a flexible work schedule f=Help with finding and getting a better job g=Any special equipment or medical devices (SPECIFY: <OPEN>) h=Is there anything else that I didn't mention that would have helped {you/NAME} to work or earn more during 2004 (SPECIFY: <OPEN>)	i=Better health/treatment
E43	Why {are you/is NAME} no longer receiving services from {EN IN 2004 FROM E39}?	<OPEN>	01=Never received any info 02=Found a job 03=Cannot work for health reasons 04=Other reason related to personal circumstance 05=Other reason related to EN 06=Other
F14	Why didn't {you/NAME or his/her representative} try to use {your/NAME's} Ticket with the State VR agency in 2004?	<OPEN>	01=Agency didn't help 02=Did not know could 03=Was not healthy enough 04=Other

Item	Question Text	Questionnaire Response Options	Categories Added During Coding
F29	After receiving information about the Employment Networks in {your/NAME's} area including the State VR agency or {STATE NAME FOR VR}, why didn't {you/NAME or his/her representative} contact any of them?	01=PHYSICAL/MENTAL CONDITION 02=CHANGED MIND 03=FAMILY RESPONSIBILITIES 04=FAMILY WOULD NOT SUPPORT 05=COULD NOT GET RELIABLE TRANSPORTATION 06=ECONOMIC CONDITIONS CHANGED 07=FEARED SERVICES WOULD ENDANGER BENEFITS 08=INFORMATION TOO CONFUSING 09=EMPLOYMENT NETWORK {NAME} WANTED WAS NOT PARTICIPATING 10=ENS TOO FAR AWAY 11=COULD NOT GET IN CONTACT WITH ENS 12=NO ENS PROVIDED SERVICES {NAME} NEEDS 13=NO ENS SERVE MY KIND OF DISABILITY 14=OTHER (SPECIFY: <OPEN>)	15=GOT A JOB OR IN SCHOOL
F31	What are the main reasons {you did/NAME did} not try to participate in the Ticket to Work program in 2004?	<OPEN>	01=Health Reasons 02=H had a job/in school 03=Did not know about program 04=Did not want to/did not try 05=Other 06=Cannot work, reason unspecified
G7	Thinking about {PROVIDER FROM G2}, was this place:	01=A state agency 02=A private business 03= Some other type of place (SPECIFY: <OPEN>)	04=School
G9	Was this place a:	01=A vocational rehabilitation agency 02=A welfare agency 03=A mental health agency 04=Some other state agency (SPECIFY: <OPEN>) 05=Some other type of place (SPECIFY: <OPEN>)	06=Workforce center/ employment office
G13	Thinking about {PROVIDER FROM G11}, was this place:	01=A state agency 02=A private business 03= Some other type of place (SPECIFY: <OPEN>)	04=School or college

Item	Question Text	Questionnaire Response Options	Categories Added During Coding
G18	Thinking about {NEW PROVIDER FROM G16}, was this place:	01=A clinic, 02=A hospital 03=A doctor's office, or 04=Some other type of place (SPECIFY: <OPEN>)	05=A school or college 06=A nursing home/group home 07=A government agency 08=In home care 09=A medical equipment store 10=A rehabilitation/counseling center 11=Physical therapy center
G22	Thinking about {NEW PROVIDER FROM G20}, was this place:	01=A mental health agency 02=A clinic 03=A hospital, 04=A doctor's office, or 05=Some other type of place (SPECIFY: <OPEN>)	06=Residential treatment program/facility 07=Rehab center/counseling center/day program
G45	In 2004, who paid for the services {you/NAME} received from {PROVIDER FROM G32 DE-DUPLICATED LIST IF USED IN 2004}?	01={NAME} 02=PROVIDER FROM G32 DE-DUPLICATED LIST IF USED IN 2004 03=NO ONE 04=FAMILY 05=EMPLOYMENT NETWORK 06=MEDICARE 07=MEDICAID 08=EMPLOYER 09=NON-PROFIT ORGANIZATION SERVING PEOPLE WITH DISABILITIES 10=WORKER'S COMPENSATION 11=DISABILITY INSURANCE 12=OTHER (SPECIFY: <OPEN>)	14=SCHOOL/FINANCIAL AID/ GRANT 15=STATE AGENCY/COUNTY/ GOVERNMENT
G53	Thinking only about the services {you/NAME} used in 2004, what are the main reasons {you/he/she} decided to use these services?	01=TO FIND A JOB/GET A BETTER JOB 02=TO INCREASE INCOME 03=TO IMPROVE HEALTH 04=TO IMPROVE ABILITY TO DO DAILY ACTIVITIES 05=TO AVOID A CONTINUING DISABILITY REVIEW 06=SOMEONE PRESSURED {NAME} TO PARTICIPATE 07=WANTED ACCESS TO A SPECIFIC PROGRAM/SERVICE/RESOURCE 08=OTHER (SPECIFY: <OPEN>)	09=TO BE MORE INDEPENDENT

Item	Question Text	Questionnaire Response Options	Categories Added During Coding
G55	Who pressured {you/NAME} to use these services?	01=PARENT/GUARDIAN 02=SPOUSE/PARTNER 03=OTHER FAMILY MEMBER 04=FRIEND/CO-WORKER 05=EMPLOYER/SUPERVISOR 06=STAFF OF EMPLOYMENT NETWORK 07=VOCATIONAL REHABILITATION CASE MANAGER 08=JOB COACH 09=SSA LETTER 10=SSA STAFF 11=BENEFIT SPECIALIST/BPAO 12=OTHER (SPECIFY: <OPEN>)	13=HEALTH CARE PROVIDER 14=COURT/POLICE
G56	How did {your/NAME's} (PERSONS(S) FROM G55) pressure :you/NAME" to use these services	01=SAID {NAME} WOULD LOSE DISABILITY AND/OR HEALTH INSURANCE BENEFITS 02=WOULD NOT TAKE "NO" FOR AN ANSWER 03=THREATENED TO WITHHOLD SERVICES 04=THREATENED TO TAKE AWAY OTHER SUPPORT 05=OTHER (SPECIFY: <OPEN>)	06=THREATENED HOSPITALIZATION/JAIL
G61	Why {were you/was NAME} unable to get these services?	<OPEN>	01=Not eligible/request refused 02=Lack of information 03=Could not afford/insurance did not cover 04=Did not try 05=Too difficult/too confusing 06=Problems with the service or agency 07=Other
H3	Why did {you/NAME} decide to participate in the Ticket to Work program?	<OPEN>	01=Wanted to get a job/ more money/benefits 02=Wanted to feel more independent 03=Other
H23	Why didn't {you/NAME or his/her representative} try to use {your/NAME's} Ticket with the State VR agency in 2004?	<OPEN>	01=Signed up with other agency 02=Already receiving services 03=Other

Item	Question Text	Questionnaire Response Options	Categories Added During Coding
H29	Why didn't {you/NAME or (his/her) representative} try to use {your/NAME's} Ticket with {any of} the other Employment Network(s) {you/NAME or (his/her) representative} contacted in 2004?	<OPEN>	01=Location 02=Other
H31	Why didn't {any of} the other { Employment Network(s) {you/NAME} tried to use {your/his/her} Ticket with accept {your/NAME's} Ticket in 2004?	01=NOT TAKING TICKETS WHEN CONTACTED 02=DID NOT OFFER SERVICES {NAME} NEEDED 03=DID NOT SERVE PEOPLE WITH {NAME'S} DISABILITY/NEEDS 04={NAME} NOT WILLING/ABLE TO WORK FULL-TIME/ENOUGH HOURS 05={NAME} NOT WILLING TO GO OFF OF DISABILITY BENEFITS 06= OTHER (SPECIFY: <OPEN>)	07=TROUBLE CONTACTING EN
H33	What information did {you/NAME} need but didn't get?	<OPEN>	01=How/where to use the ticket 02=Services provided 03=Other
H35	Why did {you/NAME or (his/her) representative} choose {{LONGEST} EMPLOYMENT NETWORK IN 2004}?	01=STAFF WERE MOST RESPONSIVE/COURTEOUS/KNOWLEDGEABLE 02=MOST WILLING TO PROVIDE THE SERVICES {NAME} WANTED 03=SERVED PEOPLE WITH {NAME'S} DISABILITY/NEEDS 04=WAIT FOR SERVICES WAS NOT TOO LONG 05=ONLY PROVIDER NEARBY/CLOSEST PROVIDER 06=ONLY PROVIDER WILLING TO ACCEPT TICKET 07=OTHER (SPECIFY: <OPEN>)	08=KNEW ABOUT THEM OR REFERRED TO THEM 09=FINANCIAL COMPENSATION
H38	What problems did {you/NAME} have during 2004 (with the services you received from EN)?	<OPEN>	01=Trouble making/keeping contact 02=Problems receiving services 03=Problems with counselor 04=Other

Item	Question Text	Questionnaire Response Options	Categories Added During Coding
H48	What was the problem about?	<OPEN>	01=Problems making/keeping contact 02=Problems receiving services 03=Other
I20	What devices, equipment, or other types of assistance {do you/does NAME} use? Anything else?	01=TELESCOPIC LENSES 02=ADAPTED COMPUTER EQUIPMENT 03=BRAILLE 04=READERS 05=GUIDE DOG 06=WHITE CANE 07=OTHER SEEING ASSISTANCE (SPECIFY: <OPEN>)	08=MAGNIFYING GLASS
I32	What devices, equipment, or other types of assistance {do you/does NAME} use? Anything else?	01=BRACES, CRUTCHES, CANE, OR WALKER 02=WHEELCHAIR OR SCOOTER 03=PROSTHETIC DEVICE 04=SPECIAL CHAIR (NOT WHEELCHAIR) 05=VEHICLE HAND CONTROLS 06=LIFT (HOME OR VEHICLE) 07=OTHER MOBILITY ASSISTANCE (SPECIFY: <OPEN>)	09=SPECIAL SHOES OR SHOE INSERTS
J11	Now, I'd like you to think back to 2004. In 2004, what kinds of health coverage did {you/NAME} have?	01=MEDICAID/{STATMED} 02=MEDICARE 03=CHAMPUS/CHAMP-VA, TRICARE, VA, OTHER MILITARY 04=INDIAN HEALTH SERVICE 05=MEDI-GAP 06=STATE PROGRAM 07=PRIVATE INSURANCE THROUGH OWN EMPLOYER 08=PRIVATE INSURANCE THROUGH SPOUSE/PARTNER/PARENT 09=PRIVATE INSURANCE PAID BY SELF/FAMILY 10=OTHER PLAN (SPECIFY: <OPEN>)	11=PRIVATE INSURANCE, NOT SPECIFIED WHO THROUGH
K14	What other assistance did {you/NAME} receive last month?	<OPEN>	01=Housing Assistance 02=Energy Assistance 03=Food assistance 04=Other

Item	Question Text	Questionnaire Response Options	Categories Added During Coding
M8	How is that person related to {you/NAME}, if at all?	01={NAME'S} SPOUSE 02={NAME'S} MOTHER 03={NAME'S} FATHER 04={NAME'S} CHILD 05=GRANDPARENT OF {NAME} 06=BROTHER/SISTER (NATURAL/STEP) OF {NAME} 07=AUNT/UNCLE OF {NAME} 08=OTHER RELATIVE OF {NAME} (SPECIFY: <OPEN>) 09=NOT RELATED (SPECIFY: <OPEN>) 10=STAFF AT RESIDENCE	11=FRIEND 12=CASEWORKER/CAREGIVER/ PAYEE 13=GIRLFRIEND/BOYFRIEND/ PARTNER 14=GUARDIAN/FOSTER PARENT/STEP PARENT 15=IN-LAW
M10	How is that person related to {you/NAME}, if at all?	01={NAME'S} SPOUSE 02={NAME'S} MOTHER 03={NAME'S} FATHER 04={NAME'S} CHILD 05=GRANDPARENT OF {NAME} 06=BROTHER/SISTER (NATURAL/STEP) OF {NAME} 07=AUNT/UNCLE OF {NAME} 08=OTHER RELATIVE OF {NAME} (SPECIFY: <OPEN>) 09=NOT RELATED (SPECIFY: <OPEN>) 10=STAFF AT RESIDENCE	11=FRIEND 12=CASEWORKER/CAREGIVER/ PAYEE 13=GIRLFRIEND/BOYFRIEND/ PARTNER 14=GUARDIAN/FOSTER PARENT/STEP PARENT 15=IN-LAW
M13	How is the assistant/proxy related to (NAME)?	01={NAME'S} SPOUSE 02={NAME'S} MOTHER 03={NAME'S} FATHER 04={NAME'S} CHILD 05=GRANDPARENT OF {NAME} 06=BROTHER/SISTER (NATURAL/STEP) OF {NAME} 07=AUNT/UNCLE OF {NAME} 08=OTHER RELATIVE OF {NAME} (SPECIFY: <OPEN>) 09=NOT RELATED (SPECIFY: <OPEN>) 10=STAFF AT RESIDENCE	11=FRIEND 12=CASEWORKER/CAREGIVER/ PAYEE 13=GIRLFRIEND/BOYFRIEND/ PARTNER 14=GUARDIAN/FOSTER PARENT/STEP PARENT 15=IN-LAW

APPENDIX B

SOC MAJOR AND MINOR OCCUPATION CLASSIFICATIONS

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

SOC MAJOR AND MINOR OCCUPATION CLASSIFICATIONS

Code	Occupation
	Management
111	Top Executives
112	Advertising, Marketing, PR, Sales
113	Operations Specialist Managers
119	Other Management Occupations
	Business /Financial Operations
131	Business Operations Specialist
132	Financial Specialist
	Computer and Mathematical Science
151	Computer Specialist
152	Mathematical Science Occupations
	Architecture and Engineering
171	Architects, Surveyors and Cartographers
172	Engineers
173	Drafters, Engineering and Mapping Technicians
	Life, Physical and Social Science
191	Life Scientists
192	Physical Scientists
193	Social Scientists and Related Workers
194	Life, Physical and Social Science Technicians
	Community and Social Services
211	Counselors, Social Workers and Other Community and Social Service Specialists
212	Religious Workers
	Legal
231	Lawyers, Judges and Related Workers
232	Legal Support Workers
	Education, Training and Library
251	Postsecondary Teachers
252	Primary, Secondary and Special Education School Teachers
253	Other Teachers and Instructors
254	Librarians, Curators and Archivists
259	Other Education, Training and Library Occupations
	Arts, Design, Entertainment, Sports and Media
271	Art and Design Workers
272	Entertainers and Performers, Sports and Related Workers
273	Media and Communication Workers
274	Media and Communication Equipment Workers
	Healthcare Practitioner and Technical Occupations
291	Health Diagnosing and Treating Practitioners
292	Health Technologists and Technicians
299	Other Healthcare Practitioner and Technical Occupations

Code	Occupation
	Healthcare Support
311	Nursing, Psychiatric and Home Health Aides
312	Occupational and Physical Therapist Assistants and Aides
319	Other Healthcare Support Occupations
	Protective Service
331	Supervisors, Protective Service Workers
332	Firefighting and Prevention Workers
333	Law Enforcement Workers
339	Other Protective Service Workers
	Food Preparation and Serving Related
351	Supervisors, Food Preparation and Food Serving Workers
352	Cooks and Food Preparation Workers
353	Food and Beverage Serving Workers
359	Other Food Preparation and Serving Related Workers
	Building and Grounds Cleaning and Maintenance
371	Supervisors, Building and Grounds Cleaning and Maintenance Workers
372	Building Cleaning and Pest Control Workers
373	Grounds Maintenance Workers
	Personal Care and Service Occupations
391	Supervisors, Personal Care and Service Workers
392	Animal Care and Service Workers
393	Entertainment Attendants and Related Workers
394	Funeral Service Workers
395	Personal Appearance Workers
396	Transportation, Tourism, and Lodging Attendants
399	Other Personal Care and Service Workers
	Sales and Related Occupations
411	Supervisors, Sales Workers
412	Retail Sales Workers
413	Sales Representative, Services
414	Sales Representative, Wholesale and Manufacturing
419	Other Sales and Related Workers
	Office and Administrative Support
431	Supervisors, Office and Administrative Support Workers
432	Communications Equipment Operators
433	Financial Clerks
434	Information and Record Clerks
435	Material Recording, Scheduling Dispatching, and Distribution Workers
436	Secretaries and Administrative Assistants
439	Other Office and Administrative Support Workers
	Farming, Fishing and Forestry Workers
451	Supervisors, Farming, Fishing and Forestry Workers
452	Agricultural Workers
453	Fishing and Hunting Workers
454	Forest, Conservation and Logging Workers

Code	Occupation
	Construction and Extraction Occupations
471	Supervisors, Construction and Extraction Workers
472	Construction Trade Workers
473	Helpers, Construction Trades
474	Other Construction and Related Workers
475	Extraction Workers
	Installation, Maintenance and Repair Occupations
491	Supervisors, Installation, Maintenance and Repair Workers
492	Electrical and Electronic Equipment Mechanics, Installers and Repairers
493	Vehicle and Mobile Equipment Mechanics, Installers and Repairers
494	Other Installation, Maintenance and Repair Occupations
	Production Occupations
511	Supervisors, Production Workers
512	Assemblers and Fabricators
513	Food Processing Workers
514	Metal Workers and Plastic Workers
515	Printing Workers
516	Textile, Apparel, and Furnishing Workers
517	Woodworkers
518	Plant and System Operators
519	Other Production Occupations
	Transportation and Material Moving Occupations
531	Supervisors, Transportation and Material Moving Workers
532	Air Transportation Workers
533	Motor Vehicle Operators
534	Rail Transportation Workers
535	Water Transportation Workers
536	Other Transportation Workers
537	Material Moving Workers
	Military Specific Occupations
551	Military Officer and Tactical Operations Leaders/Managers
552	First-Line Enlisted Military Supervisors/Managers
553	Military Enlisted Tactical Operations and Air/Weapons Specialists and Crew Members

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

APPENDIX C

NAICS INDUSTRY CODES

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

NAICS INDUSTRY CODES

Code	Description
11	**Agriculture, Forestry Fishing and Hunting**
111	Crop Production
112	Animal Production
113	Forestry and Logging
114	Fishing, Hunting and Trapping
115	Support Activities for Agriculture and Forestry
21	**Mining**
211	Oil and Gas Extraction
212	Mining (except Oil and Gas)
213	Support Activities for Mining
22	**Utilities**
221	Utilities
23	**Construction**
236	Construction of Buildings
237	Heavy and Civil Engineering Construction
238	Specialty Trade Contractors
31-33	**Manufacturing**
311	Food Manufacturing
312	Beverage and Tobacco Product Manufacturing
313	Textile Mills
314	Textile Product Mills
315	Apparel Manufacturing
316	Leather and Allied Product Manufacturing
321	Wood Product Manufacturing
322	Paper Manufacturing
323	Printing and Related Support Activities
324	Petroleum and Coal Products Manufacturing
325	Chemical Manufacturing
326	Plastics and Rubber Products Manufacturing
327	Nonmetallic Mineral Product Manufacturing
331	Primary Metal Manufacturing
332	Fabricated Metal Products Manufacturing
333	Machinery Manufacturing
334	Computer and Electronic Product Manufacturing
335	Electrical Equipment, Appliance and Component Manufacturing
336	Transportation Equipment Manufacturing
337	Furniture and Related Product Manufacturing
339	Miscellaneous Manufacturing

Code	Description
42	**Wholesale Trade**
423	Merchant Wholesalers, Durable Goods
424	Merchant Wholesalers, Nondurable Goods
425	Wholesale Electronic Markets and Agents and Brokers
44-45	**Retail Trade**
442	Furniture and Home Furnishings Stores
443	Electronics and Appliance Stores
444	Building Material and Garden Equipment and Supplies Dealers
445	Food and Beverage Stores
446	Health and Personal Care Stores
447	Gasoline Stations
448	Clothing and Clothing Accessories Stores
451	Sporting Goods, Hobby, Book, and Music Stores
452	General Merchandise Stores
453	Miscellaneous Store Retailers
454	Nonstore Retailers
48-49	**Transportation and Warehousing**
481	Air Transportation
482	Rail Transportation
483	Water Transportation
484	Truck Transportation
485	Transit and Ground Passenger Transportation
486	Pipeline Transportation
487	Scenic and Sightseeing Transportation
488	Support Activities for Transportation
491	Postal Service
492	Couriers and Messengers
493	Warehousing and Storage
51	**Information**
511	Publishing Industries (except Internet)
512	Motion Picture and Sound Recording Industries
515	Broadcasting (except Internet)
516	Internet Publishing and Broadcasting
517	Telecommunications
518	Internet Service Providers, Web Search Portals, and Data Processing Services
519	Other Information Services
52	**Finance and Insurance**
522	Credit Intermediation and Related Activities
523	Securities, Commodity Contracts, and Other Financial Investments and Related Activities
524	Insurance Carriers and Related Activities
525	Funds, Trusts, and Other Financial Vehicles

Code	Description
53	**Real Estate and Rental and Leasing**
531	Real Estate
532	Rental and Leasing Services
533	Lessors of Nonfinancial Intangible Assets (except Copyrighted Works)
54	**Professional, Scientific, and Technical Services**
55	**Management of Companies and Enterprises**
551	Management of Companies and Enterprises
56	**Administrative and Supportive Waste Management and Remediation Services**
561	Administrative and Support Services
562	Waste Management and Remediation Services
61	**Educational Services**
611	Educational Services
62	**Health Care and Social Assistance**
621	Ambulatory Health Care Services
622	Hospitals
623	Nursing and Residential Care Facilities
624	Social Assistance
71	**Arts, Entertainment, and Recreation**
711	Performing Arts Companies
712	Museums, Historical Sites, and Similar Institutions
713	Amusement, Gambling, and Recreation Industries
72	**Accommodation and Food Services**
721	Accommodation
722	Food Services and Drinking Places
81	**Other Services (except Public Administration)**
811	Repair and Maintenance
812	Personal and Laundry Services
813	Religious, Grantmaking, Civic, Professional, and Similar Organizations
814	Private Households
92	**Public Administration**
921	Executive, Legislative, and Other General Government Support
922	Justice, Public Order, and Safety Activities
923	Administration of Human Resources Programs
924	Administration of Environmental Quality
925	Administration of Housing Programs, Urban Planning, and Community Development
926	Administration of Economic Programs
927	Space Research and Technology
928	National Security and International Affairs

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

APPENDIX D

PARAMETER ESTIMATES AND STANDARD ERRORS FOR NONRESPONSE MODELS

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

LOCATION LOGISTIC PROPENSITY MODEL: REPRESENTATIVE BENEFICIARY SAMPLE

Factors in the Location Model Main Effects	Parameter Estimate	Standard Error
Number of moves in past 5 years (MOVE):		
No moves, or old information	0.054	0.723
One or two moves	Ref. cell	
No information about moves	0.837	0.438
Race (RACE):		
Non-Hispanic white, non-hispanic Asian-Pacific islander, non-Hispanic American Indian	-0.646	0.270
Hispanic, non-Hispanic Black, and Other/Unknown	Ref. cell	
Identity of payee relative to beneficiary (REPREPAYEE):		
Beneficiary received benefit payments himself/herself	0.455	0.677
Family member received benefits on behalf of beneficiary	-0.471	0.703
Institution received benefits on behalf of beneficiary	0.232	0.406
Information about who is the payee not given	Ref. cell	
Indicator whether beneficiary and applicant for benefits are in same zip code (PDZIPSAME):		
Applicant and beneficiary live in same zip code	Ref. cell	
Applicant and beneficiary live in different zip code, or no information	-0.813	0.715
Gender (SEX):		
Female	Ref. cell	
Male, or information not given	0.014	0.417
Indicator whether beneficiary resides in place with 1 million or more residents (METRO_1M):		
Beneficiary resides in place with 1 million or more residents	Ref. cell	
Beneficiary does not live in such a place	-0.003	0.161
Two-factor Interactions[a]		
MOVE * RACE		
No information about moves * Non-Hispanic White, non-Hispanic Asian-Pacific Islander, non-Hispanic American Indian	1.016	0.372
MOVE * SEX		
No moves, or old information about moves * Male or Unknown	-0.813	0.459
No information about moves * Male or Unknown	-0.142	0.478
MOVE * REPREPAYEE		
No moves, or old information about moves * Beneficiary received benefit payments himself/herself	0.375	0.716
No moves, or old information about moves * Family member received benefits on behalf of beneficiary	0.070	0.577
No moves, or old information about moves * Institution received benefits on behalf of beneficiary	1.150	0.600
MOVE * PDZIPSAME		
No moves, or old information about moves * Applicant and beneficiary live in different zip code, or no information	0.844	0.743

Factors in the Location Model Main Effects	Parameter Estimate	Standard Error
PDZIPSAME * REPREPAYEE		
Applicant and beneficiary live in different zip code, or no information * Beneficiary received benefit payments himself/herself	0.389	0.686
Applicant and beneficiary live in different zip code, or no information * Family member received benefits on behalf of beneficiary	0.956	0.717
SEX * REPREPAYEE		
Male or unknown * Beneficiary received benefit payments himself/herself	-0.430	0.331
RACE * METRO_1M		
Non-Hispanic white, non-hispanic Asian-Pacific islander, non-Hispanic American Indian * Sample member resides in place with 1 million or more residents	0.508	0.237
RACE * SEX		
Non-Hispanic white, non-hispanic Asian-Pacific islander, non-Hispanic American Indian * Male, or Unknown	0.690	0.261
Three-factor Interactions[a]		
MOVE * RACE * SEX		
No information about moves * Non-Hispanic white, non-hispanic Asian-Pacific islander, non-Hispanic American Indian * Male or Unknown	-0.733	0.403
MOVE * SEX * REPREPAYEE		
No moves, or old information about moves * Male or unknown * Beneficiary received benefit payments himself/herself	0.512	0.449
MOVE * REPREPAYEE * PDZIPSAME		
No moves, or old information about moves * Beneficiary received benefit payments himself/herself * Applicant and beneficiary live in different zip code, or no information	-1.511	0.736
No moves, or old information about moves * Family member received benefits on behalf of beneficiary * Applicant and beneficiary live in different zip code, or no information	1.448	0.779

COOPERATION LOGISTIC PROPENSITY MODEL: REPRESENTATIVE BENEFICIARY SAMPLE

Factors in the Location Model Main Effects	Parameter Estimate	Standard Error
Beneficiary residence (PSU):		
PSU1 (PSUs 4010, 6019, 6039, 6651, 6667, 12031, 12038, 17618, 25010, 28019, 36003, 36025, 36037, 37003, 47030, 53011)	0.402	0.271
PSU2 (PSUs 6037, 6040, 6650, 26021, 27002, 34019, 39027, 39031, 39049, 40024, 45012, 51019, 53003)	-0.269	0.106
PSU3 (PSUs 6686, 25009, 36035, 36040)	-0.611	0.302
PSU5 (PSUs 1026, 10002, 12025, 12027, 13045, 19004, 24009, 24016, 26028, 28022, 29035, 31009, 41007, 51039	0.225	0.097
PSU6 (PSUs 5034, 17042, 18007, 21040, 21049, 48009, 48018, 48065, 54010, 54027, 55018)	0.387	0.170
PSU4 (PSUs not identified in PSU1, PSU2, PSU3, PSU5, and PSU6)	Ref. cell	
Beneficiary recipient benefit type (SSI_SSDI):		
Beneficiary received SSI only	Ref. cell	
Beneficiary received SSDI only	-0.180	0.130
Beneficiary received both SSI and SSDI	0.320	0.201
Identity of payee relative to beneficiary (REPREPAYEE):		
Institution received benefits on behalf of beneficiary	0.260	0.179
All other payees (including those with unknown payee identity)	Ref. cell	
Indicator whether beneficiary and applicant for benefits are in same zip code (PDZIPSAME):		
Applicant and beneficiary live in same zip code	Ref. cell	
Applicant and beneficiary live in different zip code, or no information	0.027	0.237
Gender (SEX):		
Female	Ref. cell	
Male, or information not given	-0.155	0.081
Disability diagnosis classification (DIG):		
Beneficiary has mental or physical disability (excluding deaf sample members)	Ref. cell	
Beneficiary is deaf, or information about disability not given	-0.849	0.435
Two-factor Interactions[a]		
PSU1 * PDZIPSAME		
PSUs 4010, 6019, 6039, 6651, 6667, 12031, 12038, 17618, 25010, 28019, 36003, 36025, 36037, 37003, 47030, 53011 * Applicant and beneficiary live in different zip code, or no information	-1.075	0.376
PSU1 * DIG		
PSUs 4010, 6019, 6039, 6651, 6667, 12031, 12038, 17618, 25010, 28019, 36003, 36025, 36037, 37003, 47030, 53011 * Beneficiary is deaf, or no disability information is given	-0.606	0.338
PSU1 * SSI_SSDI		
PSUs 4010, 6019, 6039, 6651, 6667, 12031, 12038, 17618, 25010, 28019, 36003, 36025, 36037, 37003, 47030, 53011 * Beneficiary received both SSI and SSDI	-0.877	0.312

Factors in the Location Model Main Effects	Parameter Estimate	Standard Error
PSU2 * REPREPAYEE		
PSUs 6037, 6040, 6650, 26021, 27002, 34019, 39027, 39031, 39049, 40024, 45012, 51019, 53003 * Institution received benefits on behalf of beneficiary	-0.866	0.276
PSU3 * SEX		
PSUs 6686, 25009, 36035, 36040 * Male or Unknown	-0.653	0.289
PSU6 * SSI_SSDI		
PSUs 5034, 17042, 18007, 21040, 21049, 48009, 48018, 48065, 54010, 54027, 55018 * Beneficiary received SSDI only	0.596	0.214
SSI_SSDI * DIG		
Beneficiary received both SSI and SSDI * Beneficiary is deaf, or disability unknown	0.870	0.429
SSI_SSDI * PDZIPSAME		
Beneficiary received both SSI and SSDI * Applicant and beneficiary live in different zip code, or no information	-0.453	0.274
Three-factor Interactions[a]		
SSI_SSDI * PSU1 * PDZIPSAME		
Beneficiary received both SSI and SSDI * PSUs 4010, 6019, 6039, 6651, 6667, 12031, 12038, 17618, 25010, 28019, 36003, 36025, 36037, 37003, 47030, 53011 * Applicant and beneficiary live in different zip code, or no information	1.042	0.444

LOGISTIC PROPENSITY MODEL: TICKET PARTICIPANT SAMPLE

Factors in the Location Model Main Effects	Parameter Estimate	Standard Error
Stratum payment type (STRATA):		
Traditional or Milestones payment type	-1.286	0.662
Outcomes payment type	Ref. cell	
Region of country (REGION):		
Beneficiary resides in Northeast, South, or West region of country	-0.193	0.329
Beneficiary resides in Midwest region of country	Ref. cell	
Gender (SEX):		
Female	Ref. cell	
Male, or information not given	-0.367	0.653
Two-factor Interactions[a]		
STRATUM*REGION		
Traditional or Milestones payment type * Beneficiary resides in Northeast, South, or West region of country	0.422	0.623
STRATUM*SEX		
Traditional or Milestones payment type * Male or Unknown	0.810	0.750

[a] All combinations for the listed interactions that are not shown are part of the reference cells.

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

APPENDIX E

SUDAAN AND SAS PARAMETERS USED TO OBTAIN NATIONAL ESTIMATES

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

SUDAAN PARAMETERS

SUDAAN EXAMPLE

```
proc descript data="SASdatasetname" filetype=sas design=wr;
nest        A_STRATA A_PSU / missunit;
weight      "weight variable" ;
subpop      "response variable" = "complete";
var         "analysis variables" ;
print nsum wsum mean semean deffmean / style=nchs
wsumfmt=f10.0 meanfmt=f8.4 semeanfmt=f8.4 deffmeanfmt=f8.4;
title       "TTW National Estimates";
```

SAS EXAMPLE

```
proc surveymeans data="SASdatasetname";
strata   A_STRATA;
cluster  A_PSU;
weight "weight variable" ;

where  "response variable" = "complete";
var      "analysis variables" ;
title    "TTW National Estimates";
```

Weight Variables

Beneficiary sample:	**Wgt1_Benefinl**
Participant sample:	**Wgt1_Partifinl**
Combined samples:	**Wgt1_Combfinl**

Nest Variables

A_STRATA

1. Clustered samples for both beneficiaries and participants
   a. A_STRATA = 100 for PSUs in Phase 1 states
   b. A_STRATA = 200 for PSUs in Phase 2 states
   c. A_STRATA = 300 for PSUs in Phase 3 states

2. Unclustered samples for participants requiring unclustered sample
   a. A_STRATA = 111 Outcome-only participants in PSUs in Phase 1 states
   b. A_STRATA = 112 Outcome-only participants not in PSUs in Phase 1 states

A_PSU

1. Clustered samples for both beneficiaries and participants
   A_PSU = PSU identifier
2. Unclustered samples for participants requiring unclustered sample
   A_PSU = MPR_ID for Outcome-only participants

Notes:

1. Before each SUDAAN procedure, sort by A_STRATA and A_PSU
2. Use SUDAAN's SUBPOP statement to define population for which estimates are wanted. For example, for estimates of SSI participant population, use SUBPOP to define SSI participants.