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

MATHEMATICA
Policy Research, Inc.

National Beneficiary Survey Round 2 (Volume 1 of 3): Editing, Coding, Imputation, and Weighting Procedures

December 19, 2008

Eric Grau
Debra Wright
Yuhong Zheng
Frank Potter

Submitted to:

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

Project Officer:
Paul O'Leary

Submitted by:

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

Project Director:
Debra Wright

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW 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**

TABLES

TABLES *(continued)*

FIGURES

Figure **Page**

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

ACRONYMS

AIC: Akaike's Information Criterion

CAPI: Computer-assisted personal interviewing

CATI: Computer-assisted telephone interviewing

CHAID: Chi-Square Automatic Interaction Detector

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 is a registered trademark of SAS Institute, Inc., Cary, NC)

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 StataCorp LP, College Station, TX.)

TTW: Ticket to Work

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

I. INTRODUCTION

As part of an evaluation of the Ticket to Work and Self-Sufficiency program (TTW), Mathematica Policy Research, Inc. (MPR) conducted the second round of the National Beneficiary Survey (NBS) in 2005. 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 (hereafter referred to as the Representative Beneficiary Sample) and a sample of TTW participants (hereafter referred to as the Ticket Participant Sample). The Ticket Participant Sample contains cross-sectional and longitudinal components, both of which are discussed in this report. MPR collected data using computer-assisted telephone interviewing (CATI) and computer-assisted personal interview (CAPI) follow-ups of CATI nonrespondents and those who preferred or needed an in-person interview to accommodate their disability.

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 itself, SSA provides 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

[1] For more information on the Ticket to Work Program, see Thornton et al. 2004.

national cross-sectional surveys (called rounds) of Ticket-eligible SSA disability beneficiaries—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 are to be 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 completes cases is provided in Chapter III.

[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 the following 13 states, known as the "Phase 1" states: Arizona, Colorado, Delaware, Florida, Illinois, Iowa, Massachusetts, New York, Oklahoma, Oregon, South Carolina, Vermont, and Wisconsin. The Phase 2 roll-out ran from November 2002 through September 2003, during which time SSA distributed Tickets in the following 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 roll-out ran from November 2003 through September 2004, during which time SSA distributed Tickets in 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.

The NBS used a multi-stage sampling design (which was used for all survey rounds) 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 the 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 two counties with a large number of beneficiaries, 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 type of payment system under which SSA paid a service provider: the traditional vocational rehabilitation payment system, the milestone-outcome payment system, and 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 2 User's Guide (Wright, et al. 2008) contains more information on the round 2 sampling design.

In round 1 (2004), two surveys were fielded: 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). Three cross-sectional surveys were fielded in round 2 (2005):

1. The second national survey of all beneficiaries (The Representative Beneficiary Sample),
2. The second cross-sectional survey of Ticket participants who resided in a Phase 1 state at the time of Ticket assignment (The Phase 1 Cross-Sectional Ticket Participant Sample), and
3. The first cross-sectional survey of Ticket participants who resided in a Phase 2 state at the time of Ticket assignment (The Phase 2 Cross-Sectional Ticket Participant Sample).

At round 2, we also attempted to re-interview Phase 1 Ticket Participants who were selected into the sample at round 1, whether or not they had been interviewed in round 1 (the Phase 1 Longitudinal Sample). The original sample design called for re-interviewing only those longitudinal cases that had completed the previous round. However, based on MPR's recommendation, interviews were attempted with all longitudinal cases.

[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.

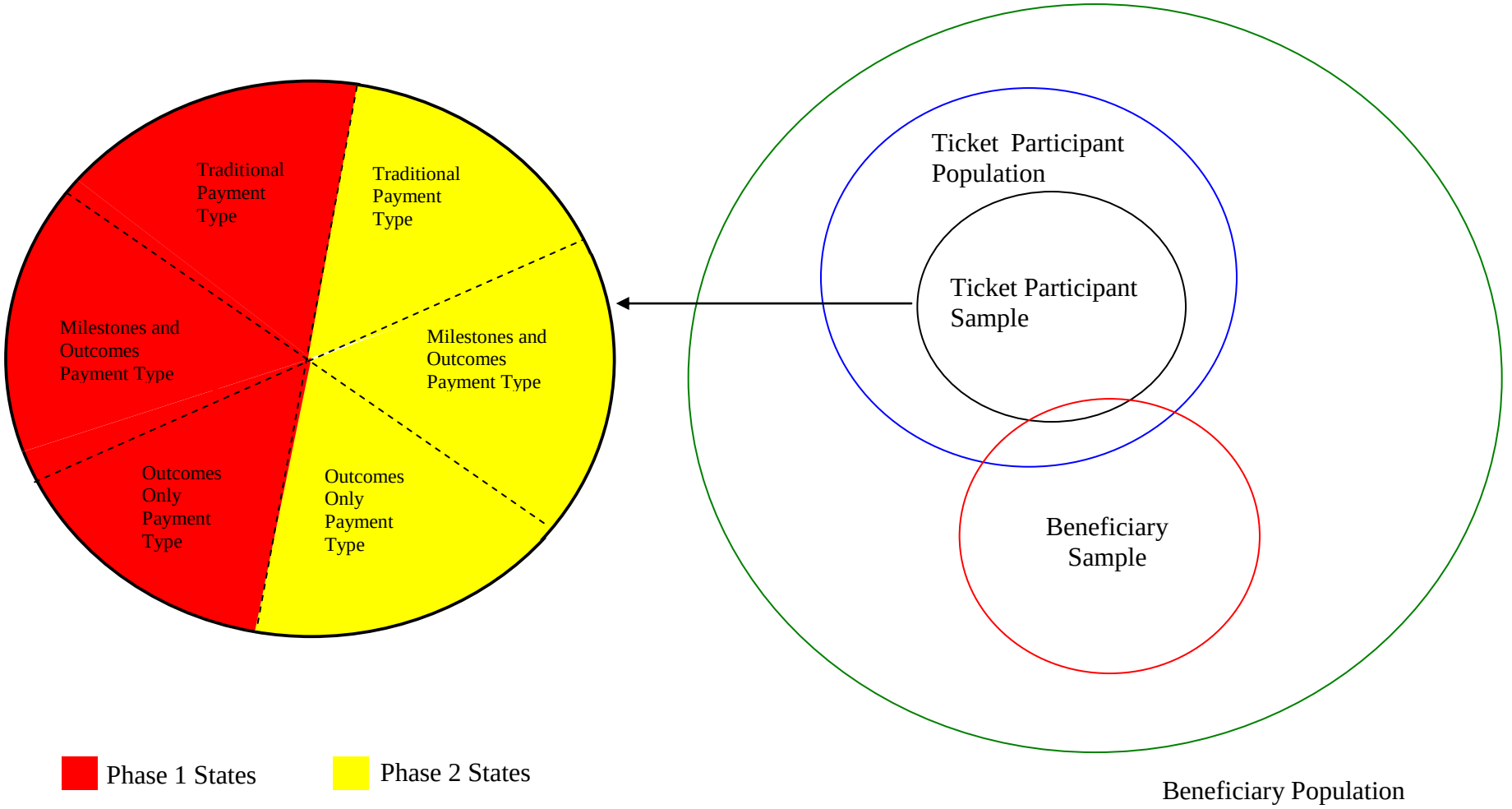
In the first follow-up year (round 2 for Phase 1 participants), a supplemental sample of those who had entered the Ticket program since the first year of rollout for each phase, or otherwise had not been sampled before, was selected to produce an expanded second-year cross-sectional sample survey. For Phase 1 participants, this resulted in cross-sectional samples for two consecutive years. The cross-sectional surveys consisted of the supplemental cases, plus the longitudinal cases who were still Ticket participants at the time of sampling.

Figure I.1 shows how the various samples and populations relate to one another.[4] The population of Ticket participants, represented by the blue circle, is a subset of the population of all Supplemental Security Income (SSI) and Social Security Disability Insurance (SSDI) beneficiaries, as represented by the green circle. The Representative Beneficiary Sample (represented by the red circle) could include some individuals who are also Ticket participants (there were 61 such cases in round 2, where the red circle and blue circle intersect). Moreover, it is possible for a Ticket participant to have been selected for both the Representative Beneficiary Sample and the Ticket Participant Sample (there were 24 such cases in round 2, where the red circle and black circle intersect). The samples taken from these populations represent a snapshot of the populations at round 2, so that the Ticket Participant Sample does not include individuals in Phase 3 states. The Ticket Participant Sample, as shown in Figure I.1, also does not include Phase 1 longitudinal sample cases who were no longer in the program in round 2, since they were no longer part of the Ticket participant population. Finally, the figure shows the Ticket participant subpopulation as a much larger proportion of the beneficiary population than is actually the case. In fact, in round 2, the Ticket participant subpopulation was less than 0.6 percent of the entire beneficiary population.

[4]The composition of the populations and samples represented by these circles changes from round to round. For example, a round 1 snapshot would show only Phase 1 cases in the Ticket Participant Sample; a round 3 snapshot would also show a Phase 3 subsample in addition to the Phase 1 and Phase 2 subsamples.

FIGURE I.1

REPRESENTATIVE BENEFICIARY AND TICKET PARTICIPANT SAMPLES AND POPULATIONS AT ROUND 2



B. NBS OBJECTIVES

The NBS is one of several components of an evaluation of the impact of TTW relative to the current system, the SSA Vocational Rehabilitation Reimbursement Program, which has been in place since 1981. The evaluation includes a process analysis as well as an impact and a 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 objectives:

1. To provide critical data on the work-related activities of SSI and SSDI beneficiaries, particularly as these activities relate to TTW implementation
2. To collect data on the characteristics and program experiences of beneficiaries who use their Ticket
3. To gather information about beneficiaries who do not use their Ticket, and the reasons for this choice
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 beneficiary perceptions about TTW and other SSA programs designed to help SSA beneficiaries with disabilities find and keep jobs

Round 2 NBS data will be combined with SSA administrative data to provide critical information on access to jobs and on employment outcomes for 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 disability, interest in work, use of services, and employment. As a result, SSA and external researchers who are interested in disability and employment issues can use the survey data for other policymaking and program-planning efforts.

C. ROUND 2 SURVEY OVERVIEW

As in round 1, round 2 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 2 User's Guide (Wright, et al. 2008). The survey instrument is available from SSA or MPR upon request.

Round 2 CATI data collection for both samples began in February 2005. Beginning in May 2005, 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 an in-person interview because of his or her disability, 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 or Spanish, interpreters conducted 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 2 sample comprised 6,712 cases selected for the Representative Beneficiary Sample and 4,555 cases for the Ticket Participant Sample (for a total of 11,267 cases).

TABLE I.2

ROUND 2 SAMPLE SIZES, TARGET COMPLETES, AND ACTUAL COMPLETES

Sampling Strata	Sample Size	Target Completes	Actual Completes
Representative Beneficiary Sample	6,712	4,800	4,864
Ticket Participant Sample	4,555	2,922	3,242
Phase 1 Longitudinal Participant Sample	1,466	922	1,019
Phase 1 Supplemental Participant Sample	1,739	1,000	1,230
Phase 2 Ticket Participant Sample	1,350	1,000	993
Total Sample Size	**11,267**	**7,722**	**8,106**

Source: NBS, round 2.

The round 2 CATI and CAPI data collection was completed in September 2005. Interviews were completed with 4,864 individuals in the Representative Beneficiary Sample and with 3,242 people in the Ticket Participant Sample for a total of 8,106 cases completed.[5] An additional 375 beneficiaries and 63 TTW participants were determined to be ineligible for the survey.[6] Across both samples, 6,371 cases were completed by telephone, and 1,735 were completed by CAPI. Proxy interviews were completed for 1,793 sample members. There were 207 cases in which the sample member was unable to participate and a proxy could not be identified. The weighted

[5] Because the clustered and unclustered samples of the Ticket Participant Sample were independent, it was not uncommon for individuals to be chosen for both samples. It was also possible for a sample member to be chosen for both the Representative Beneficiary Sample and the Ticket Participant Sample. Interviews for these duplicate cases were conducted only once but recorded twice (once for each sample). The counts given above include these duplicates as separate cases.

[6] Ineligible sample members include those who were deceased, incarcerated; those no longer living in the continental United States; and those whose benefit status was pending. For the Ticket Participant Sample, ineligibles also included sample members who left the program after sampling was completed (although those who were in the round 1 sample and subsequently left the program were eligible for the Phase 1 longitudinal sample).

response rate for the Representative Beneficiary Sample was 78.7 percent. The weighted response rates for the Ticket Participant Sample was 80.4 percent.

D. NBS DATA DOCUMENTATION REPORTS

The following reports make up the complete documentation describing the NBS, the round 2 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 2 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 2 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 and public use data files 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.

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

- ***NBS Questionnaire*.** This document contains all items on the round 2 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 2 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 2 sample.

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

II. DATA EDITING AND CODING

Prior to imputation, the NBS data were edited and coded to create an NBS data file. This chapter documents 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 errors in skip patterns. 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 cases for which 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 Barrett and Wright (2008) for more detail on the editing and cleaning procedures.)

For all items with fixed field numeric responses (such as number of weeks, number of jobs, dollar amounts, and so on), we reviewed the upper and lower values 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 spectrum 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 there appeared to be an error in data entry.

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 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 these 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 at round 1, 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). Categories developed during round 1 were applied at round 2. Additional categories were added at round 2 for a small number of items, if there were a significant number of common responses that did not fit into previously developed categories. 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 details 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 2.

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. Longitudinal cases that completed round 1 skipped items B6, B12, and B15 at round 2.

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).

As in round 1, the 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).

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. In total, approximately 15 percent of all coded responses were reviewed by a supervisor, including cases flagged by coders for review that they were unable to code or did not know how to code. Approximately 5 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.

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
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
03	Endocrine/nutritional disorders	Thyroid disorders, diabetes, abnormal growth disorders, nutritional disorders, and other metabolic and immunity disorders	240.0–279.9
04	Blood/blood-forming	Diseases of blood cells and spleen	280.0–289.9
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
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
07	Diseases of circulatory system	Heart disease, disorders of circulation, and diseases of arteries, veins, and capillaries	390-459.9
08	Diseases of respiratory system	Disorders of the nasal, sinus, upper respiratory tract, and lungs including chronic obstructive pulmonary disease	460-519.9
09	Diseases of digestive system	Diseases of the oral cavity, stomach, esophagus, and duodenum	520.0-579.9
10	Diseases of genitourinary system	Diseases of the kidneys, urinary system, genital organs, and breasts	580.0-629.9
11	Complications of pregnancy, child birth, and the puerperium	Complications related to pregnancy or delivery, and complications of the puerperium	630-677
12	Diseases of skin/ subcutaneous tissue	Infections of the skin, inflammatory conditions, and other skin diseases	680.0-709.9

TABLE II.2 *(continued)*

Code	Label	Description of ICD-9 Codes	Corresponding ICD-9 codes
13	Diseases of musculoskeletal system	Muscle, bone, and joint problems including arthropathies, dorsopathies, rheumatism, osteopathies, and acquired musculoskeletal deformities	710.0-739.9
14	Congenital anomalies	Problems arising from abnormal fetal development, including birth defects and genetic abnormalities	740.0-759.9
15	Conditions in the perinatal period	Conditions that have origin in birth period even if disorder emerges later	760.0-779.9
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
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
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
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 2.

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 to the occupation items were coded using the Bureau of Labor Statistic’s 2000 Standard

Occupational Classification (SOC).[7] 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 describe the major group the occupation belonged to 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.

As for round 1, verbatim responses to the industry items were coded according to the 2002 North American Industry Classification System (NAICS).[8] The NAICS 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

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

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

first several cases that were coded. In total, approximately 20 percent of all coded responses were reviewed by a supervisor, including cases flagged by coders for review that they were unable to code or did not know how to code. Approximately 10 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	Uncodable	A code cannot be assigned based on the verbatim response.

Source: NBS, round 2.

III. SAMPLING WEIGHTS

The final analysis weights for the Representative Beneficiary Sample and the Ticket Participant Sample were determined via a four-step process: (1) calculate the initial weights, (2) adjust weights for two phases of nonresponse (location and completion), (3) trim the weights to reduce the variance, and (4) poststratification. This chapter describes these computations for both the Representative Beneficiary Sample and the Ticket Participant Sample. Section A summarizes the procedures used to compute and adjust the sampling weights, the procedure for creating composite weights. (Composite weights were used in both round 1 and round 2 to combine the Representative Beneficiary Sample and Ticket Participant Sample, and to combine two samples in the Ticket Participant Sample.) Procedures for 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. Section D explains the procedures for variance estimation.

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.[9] We used four age-based strata in each PSU. Because we used a composite size

[9] 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.

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.[10]

For the initial beneficiary sample, we selected more individuals than we expected to need, to account for differential response and eligibility rates in both the PSUs and the 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 members. 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) to only those PSUs and strata requiring them, and were thus able to achieve achieved sample sizes that were close to our targets. Controlling the release of the sample also allowed us to control the balance between data collection costs and response rates. The initial sampling weights were computed on the basis of 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 SSA.[11]

We used logistic regression methods to estimate response propensities. This was done in two stages: (1) estimating a propensity score for locating a sample member and (2) estimating 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 factors

[10] 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).

[11] These totals were obtained from a frame file provided by SSA that contains basic demographics for all SSI and SSDI beneficiaries.

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 deliberately avoid the interviewer or may be expressing displeasure or hostility toward surveys in general or toward SSA in particular.

To develop the logistic propensity models for round 2, we used information from the SSA data files and geographic information (such as urban/rural or 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. We used a higher significance level because the purpose of the model was to improve the estimation of the propensity score, 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. Any covariate or interaction that was clearly unrelated to locating the respondent or to response propensity was excluded from the pool.

The next step was to carefully evaluate a series of models by comparing the following measures of predictive ability and goodness of fit: the R-squared statistic,[12] Akaike's Information Criterion (AIC),[13] percentage concordant and discordant,[14] and the Hosmer-

[12] 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.

[13] Akaike's Information Criterion is defined as AIC = -2LogL + 2(k+s), where LogL is the loglikelihood of the binomial distribution using the parameters from the given model, k is the total number of response levels minus one, and s is the number of explanatory effects (Akaike, 1974). AIC is a relative number, and has no meaning on its own. For a given model, smaller values of AIC are better than larger values.

[14] A pair of observations is concordant if a responding subject has a higher predicted value than the nonresponding subject, discordant if not, and tied if both members of the pair are either respondents, nonrespondents, or have the same predicted values. It is desirable to have as many concordant pairs and as few discordant pairs as is possible (Agresti 1996).

Lemeshow goodness-of-fit test.[15] Model-fitting also involved reviewing the statistical significance of the coefficients of the covariates in the model and avoiding any unusually large adjustment factors. In addition, we also avoided data warnings in SUDAAN.[16] We then used the specific covariate values for each located person (cooperating person) to estimate a propensity to be located (to cooperate), from which we calculated the adjusted weights. The location-adjusted weight is the product of the released adjusted weight and the inverse of the location propensity score; the nonresponse-adjusted weight is the product of the location-adjusted weight and the inverse of the cooperation propensity score.

Once the adjustments were made, we trimmed the survey weights (if necessary) to avoid unusually large weights, which would make the survey estimates less precise. We used the design effect attributable to the variation in the sampling weights as a statistical measure to determine both the necessity and the amount of trimming. The design effect attributable to weighting is a measure of the potential loss in precision caused by the variation in the sampling weights relative to a sample of the same size with equal weights. We also wanted to minimize the extent of trimming to avoid the potential for bias in the survey estimates. For the Representative Beneficiary Sample, no weights were trimmed.

The final step is a series of post-stratification adjustments through which the weights sum to known totals obtained from SSA on various dimensions (specifically, gender, age grouping, and

[15] The Hosmer-Lemeshow Goodness-of-Fit Test, unlike the Pearson and deviance goodness-of-fit tests, can be used to test goodness of fit even when some of the covariates are continuous (Hosmer and Lemeshow 1989).

[16] SUDAAN data warnings usually included one or more of the following: (1) an indication of a response cell with zero count; (2) one or more parameters approaching infinity (which may not be readily observable with the parameter estimates themselves); and (3) degrees of freedom for overall contrast less than the maximum number of estimable parameters. We tried to avoid all of these warnings, though avoiding the first two were of the highest priority. These warnings were almost always caused by a response cell with a count that was too small, which required dropping covariates or collapsing categories in covariates.

for beneficiaries only, recipient status[17]). After post-stratification, we checked the survey weights again to determine whether more trimming was necessary. In round 2, trimming was not necessary either before or after post-stratification in the Representative Beneficiary Sample.

2. Ticket Participant Samples

The initial sampling weights for the Ticket Participant Samples were computed on the basis of the inverse of the selection probability for the participant. As in the Representative Beneficiary Sample, we used the PSUs as the primary source of the sample members and, when possible, selected an initial larger (augmented) sample. For participants using either the milestone-outcome or the outcome-only payment system, the PSUs in the initial sampling design did not have enough participants to support analysis tasks—even with all participants in the PSUs from these two payment types selected for the sample. As a result, it was necessary to supplement the sample from the PSUs with a second independent sample of Ticket participants from two geographic strata defined by the PSUs. The sample members within the initial sample design are referred to as the clustered sample, and the secondary sample, which was randomly selected from the entire population of milestone-outcome and outcome-only participants in two geographic strata (in the PSUs and not in the PSUs) are referred to as the unclustered sample.[18]

As in the Representative Beneficiary Sample, we computed the weights for the augmented sample and then adjusted them for the number of sample members that were in the final

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

[18] Because of the small populations for the Outcome Only and Milestone and Outcome payment types, Ticket participants who resided in the selected PSUs for these payment types were often selected for both the clustered and the in-PSU strata of the unclustered samples. Hence, these duplicate cases had to be accounted for in the weighting process, as is discussed later.

sample.[19] We adjusted for nonresponse separately for locating sample members, and then for response among the located sample members. Using the general techniques that we applied in the Representative Beneficiary Sample, we fitted logistic propensity models to obtain the weight adjustments. The size of the sample for the three payment types was similar, but the size of the population for each was very different. Hence, the sampling weights differed substantially in magnitude from one payment system to the next. As a result, we conducted the weight adjustments separately for each payment type. For the subsamples associated with each phase and payment type within the Ticket Participant Sample, we trimmed the weights to ensure that the design effect due to unequal weighting was not substantially greater than 3.0 (less than 3.0 if possible). The final adjustment for the participants' weights was a post-stratification adjustment to the counts of participants within subgroups defined by age and gender in the sampling frame. After post-stratification, we checked the survey again to determine whether more trimming was necessary. In round 2, although trimming was required before post-stratification in the Ticket Participant Sample, no trimming was required after post-stratification.

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

[19] 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.)

In round 1, we used a sophisticated procedure to create these weights such that the variance of survey estimates was minimized. This procedure allowed for weights to be applied to observations that were duplicated across the two samples.[20] However, because the Ticket participants were such a small fraction of the beneficiary sample frame, we used a simpler alternative method at round 2.

The Representative Beneficiary Sample included few Ticket participants with completed interviews (or who were ineligible after the sample was selected): 48 of the 61 Ticket participants in the sample had completed interviews or were ineligible after sample selection in round 2, including 29 from Phase 1 states and 19 from Phase 2 states. We therefore assigned a value of zero to the original Representative Beneficiary Sample weights among these 48 cases. To ensure that the Ticket participant population would be represented, we replaced these members of the Representative Beneficiary Sample by the 3,156 members of the Ticket Participant Sample (2,149 from Phase 1 states and 1,005 from Phase 2 states) with completed interviews (or ineligible dispositions after sample selection).[21] The total sum of weights added up to the total number of Ticket participants (34,312 for Phase 1 states and 21,196 for Phase 2 states). Since the 48 Ticket participants in the Representative Beneficiary Sample did not have weights summing to the appropriate marginal totals for each phase, it was necessary to ratio-adjust the remaining beneficiary weights to the appropriate totals.

[20] A complex procedure was also used to combine the clustered and unclustered samples of the Ticket Participant Sample in both round 1 and round 2. This procedure is described in Section C of this chapter.

[21] This does not include sample members who were selected for the round 1 Ticket participant sample, were no longer Ticket participants in round 2, but were sampled anyway for longitudinal purposes.

4. Quality Assurance

To ensure that the methods used to compute the weights at each step were sound, a senior statistician conducted a final quality assurance check of the weights from both the Representative Beneficiary and Ticket Participant samples, as well as the composite weights. We chose a statistician who was not directly involved in the project for the sake of objectivity.

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.[22] 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.[23] 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 3,999 sample the members (three times the target sample size of 1,333); for beneficiaries 50 to 64 years, the 14,400 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.1.

[22] The sample of PSUs contained 79 unique selections. Because of the size of its beneficiary population, the PSU representing Los Angeles County (LA) 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 (83 units) are treated as PSUs for the beneficiary sample.

[23] 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.1

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

Sampling Strata (ages as of June 30, 2004)	Survey Population[a]	Augmented Sample Size	Global Sampling Rate (F_j)	Initial Sample Weights	Released Sample
Beneficiaries between 18 and 29 years old	1,012,037	3,999	0.003951	253.1	1,891
Beneficiaries between 30 and 39 years old	1,281,996	3,999	0.003119	320.6	1,837
Beneficiaries between 40 and 49 years old	2,461,591	3,999	0.001625	615.6	1,858
Beneficiaries between 50 and 64 years old	5,250,284	2,403	0.000458	2184.9	1,126
Total	**10,005,908**	**14,400**			**6,712**

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.

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 strata). 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 6,712 beneficiaries, as shown under "Released Sample" in Table III.1.

2. Nonresponse Adjustments

In essentially all surveys, the sampling weights must be adjusted to compensate for 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. Using variables obtained from SSA databases, a pool of covariates from which to choose a final location model was selected through stepwise regression. This pool included both main effects and interactions. From this pool of covariates, candidate models were compared using various measures of goodness of fit and predictive ability, while avoiding large adjustments. This process was repeated for interview respondents among the located sample members, 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 if the sample member was determined to be ineligible after sample selection (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, ethnicity, 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), number of changes in the beneficiary's phone number over the past five years, primary disability classification, type of disability claim (a person with a disability, a survivor, or other), living situation of beneficiary, source of data for address characteristics, whether the beneficiary was institutionalized, 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 given 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.[24]

b. Response Rates

The 78.7 percent response rate for the Representative Beneficiary Sample that is quoted in the introduction to this document is the **weighted overall completion rate**, given in the first line of Table III.2. 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. It can be determined by taking the product of the

[24] Disposition codes 420 (institutionalized) and 430 (unavailable during field period) were classified as nonrespondent codes in round 2, even though they were considered ineligible codes in round 1. This affected 8 cases in the round 2 beneficiary sample. As a result, the nonresponse adjusted weight for these 8 cases was 0 in round 2, even though a similar response in round 1 would have resulted in a positive weight. Because of the small numbers, the effect on response rates is very small.

weighted location rate and the weighted cooperation rate, also known as the weighted completion rate among located sample members.

TABLE III.2

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	6,712	6,157	93.0	5,239	84.6	78.7
SSI Only, SSDI Only, or Both SSI and SSDI						
SSI Only	2,893	2,628	92.3	2,271	86.4	79.8
SSDI Only	2,441	2,263	93.9	1,887	82.9	77.8
Both SSI and SSDI	1,378	1,266	91.7	1,081	86.4	78.6
SSI or SSDI						
SSI Only or in Both SSI & SSDI Programs	4,271	3,894	93.3	3,352	86.4	79.7
SSDI Only or in Both SSI & SSDI Programs	3,819	3,529	93.3	2,968	83.7	78.2
Constructed Disability Status						
Deaf	89	75	89.3	58	73.8	65.8
Mental	3,637	3,296	91.6	2,783	83.9	76.8
Physical	2,621	2,454	94.3	2,108	85.1	80.3
Unknown	365	332	92.3	290	87.3	80.4
Beneficiary's Age (Four Categories)						
18-29 Years	1,891	1,719	91.0	1,496	87.3	79.4
30-39 Years	1,837	1,680	91.7	1,407	83.9	77.0
40-49 Years	1,858	1,697	91.5	1,453	85.8	78.6
50-64 Years	1,126	1.061	94.4	883	83.7	79.1
Sex						
Male	3,430	3,136	92.21	2,637	83.4	76.9
Female	3,282	3,021	93.8	2,602	85.8	80.5
Hispanicity						
Non Hispanic	6,366	5,843	93.1	4,968	84.6	78.8
Hispanic	346	314	91.0	271	85.0	77.4
Race (Detailed)						
White	4,034	3,724	93.3	3,185	85.2	79.5
Black	1.492	1,345	92.0	1,146	84.5	77.7
Unknown	711	657	93.6	555	84.2	78.9
Asian American, Pacific Islander	94	85	94.6	53	60.3	56.7
North American Indian or Alaskan Native	35	32	92.8	29	93.3	86.2
Living Situation						
Living Alone	3,884	3,554	92.4	3,049	86.5	79.9
Living with Others	270	248	92.3	211	85.7	79.0
Living with Parents	72	58	81.1	53	91.7	74.6
In Institution or Unknown	2,486	2,297	93.8	1,926	82.8	77.6

TABLE III.2 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Did the Applicant for Benefits Live In Same Zip Code as Beneficiary?						
No	583	502	87.2	422	85.2	74.3
Yes	2,923	2,700	93.1	2,561	87.3	82.0
No Information	3,206	2,955	93.8	2,256	82.5	76.8
Identity of the Payee with Respect to the Beneficiary						
Beneficiary Received Beneficiary Payments						
Himself or Herself	4,236	3,871	93.3	3,272	84.3	78.7
Payee is a Family Member	1,834	1,712	93.5	1,478	86.3	80.6
Payee is an Institution	434	397	91.0	337	82.0	74.7
Other	208	177	86.3	152	86.9	75.1
Changes in Telephone Number						
No Changes in Last 5 Years	4,036	3,734	93.8	3,161	84.9	79.7
One Change in Last 5 Years	350	311	91.1	275	89.8	81.9
Two or More Changes in Last 5 Years	110	99	93.1	79	79.5	73.8
No Information on Phone Number	2,216	2,013	91.3	1,724	83.2	76.0
Number of Moves in Last 5 Years						
No Moves Last 5 Years	2,152	1,979	93.7	1,717	87.8	82.4
One or More Moves in Last 5 Years	351	294	86.1	252	87.8	75.4
No Information on Number of Moves	4,209	3,884	93.1	3,270	83.2	77.4
Type of Claim						
Disabled	3,223	2,986	93.7	2,503	83.7	78.4
Survivor	656	597	90.3	512	84.9	76.8
Unknown	2,833	2,574	92.4	2,224	86.3	79.8
Address of Payee Obtained from SSI File						
Yes	5860	5355	92.5	4576	85.1	78.8
No	852	802	94.7	663	82.7	78.3
Census Region						
Midwest	1,600	1,483	93.9	1,265	84.5	79.4
Northeast	1,093	995	92.1	824	82.9	76.4
South	2,781	2,576	94.1	2,237	85.6	80.6
West	1,238	1,103	90.2	913	84.0	75.8
Census Division						
East North Central	1,207	1,110	93.0	947	84.7	78.8
East South Central	574	528	93.6	454	84.1	78.6
Middle Atlantic	743	682	91.8	557	81.8	75.1
Mountain	330	294	91.2	250	86.7	79.2
New England	350	313	92.8	267	85.6	79.5
Pacific	908	809	89.8	663	83.0	74.5
South Atlantic	1,537	1,425	94.6	1,227	84.6	80.0
West North Central	393	373	96.6	318	83.9	80.9
West South Central	670	623	93.4	556	89.4	83.5
MSA / PMSA Size						
Not an MSA / PMSA	1,321	1,248	95.2	1,083	86.5	82.4
Areas of 1 million or more	2,902	2.631	92.1	2,206	83.2	76.6
Areas under 1 million	2,489	2,278	92.9	1,950	85.2	79.2

TABLE III.2 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Rural/Urban Continuum Code						
Metropolitan Areas of 1 Million Population or More	2,902	2,631	92.1	2,206	83.2	76.6
Metropolitan Areas of 250,000 to 999,999 Population	1,724	1,574	93.0	1,343	85.8	79.8
Metropolitan Areas of less than 250,000 Population	765	704	92.7	607	83.8	77.7
Nonmetropolitan areas adjacent to large metropolitan areas	477	452	96.1	387	81.5	78.4
Nonmetropolitan areas adjacent to medium or small metropolitan areas	517	486	93.4	423	87.4	81.7
Nonmetropolitan areas not adjacent to metropolitan areas	327	310	96.6	273	92.3	89.1

Source: NBS, round 2.

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.2 as 93.0 percent. The **weighted cooperation rate** (the weighted completion rate among located sample members), 84.6 percent in Table III.2, 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.[25] Weighted cooperation rates reflect the common survey situation that once a person is located, repeated contact efforts often will result in a completed interview.

[25] 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.

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.1, 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.2 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 sample members for whom a completed interview was obtained or who were 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.

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 has a 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 interactions among these variables for the modeling, we first ran a chi-squared automatic interaction detector (CHAID) analysis in SPSS to find possible significant interactions. 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). CHAID was first run with all covariates, then rerun a few times with the top variable in the tree removed, to ensure all potentially important interactions were retained for further consideration. The resulting pool of covariates was further reduced by evaluating tabulations of all the main effects and the interactions identified by CHAID. At a particular level of a given covariate or

interaction, if all respondents were either located or unlocated (for the location models), complete or not complete (for the cooperation models), or the total number of sample members at that level was fewer than 20, then levels were collapsed if collapsing was possible. If collapsing was not possible, then the covariate or interaction was excluded from the pool.[26]

All the resulting candidate main effects and the 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. [27] 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 the stepwise procedures, we included variables or interactions that had a statistical significance level (alpha level) of 0.30 or lower (instead of the standard 0.05).[28] Once the candidate list of main effects and interactions was determined, a thorough model-fitting process was used to determine a parsimonious model with few very small propensities. Model selection criteria were described in the overview of this chapter (Section A). The variables used as main effects and the interactions in the model are summarized in Table III.3 for locating a sample member and in Table III.4 for cooperation among located sample

[26] Deafness has historically been shown to be an important indicator of both locating a sample member, and of whether the sample member completed the interview. For that reason, deafness was allowed to remain in the covariate pool even though the number of deaf cases was sometimes as low as 18.

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

[28] As stated earlier, we used a higher significance level because the purpose of the model was to improve the 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.

members. The R-squared is 0.028 (0.071 when rescaled to have a maximum of 1) for the location model and 0.049 (0.085 when rescaled) 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. For the location model, the proportion of concordant pairs is 63 percent, 35.7 percent of the pairs are discordant, and the p-value for the chi-square statistic from the Hosmer-Lemeshow (H-L) goodness-of-fit test is 0.272; these values indicate a reasonably good fit of the model to the data. For the cooperation model, the proportion of concordant pairs is 59.6 percent, and 39.5 percent of pairs are discordant. The p-value for the chi-square statistic for the (H-L) goodness-of-fit test is 0.868 for this model. Since the AIC is a relative number, and has no meaning on its own, values for the AIC are not provided here.

TABLE III.3

LOCATION LOGISTIC PROPENSITY MODEL: REPRESENTATIVE BENEFICIARY SAMPLE

Factors in the Location Model
Main Effects
MOVE_1
DIG_1
REPREPAYEE_1
PDZIPSAME
SEX
METRO_1
INSTIT
DIVISION
SSIADDP
LIVING_1
PHONE_1
AGECAT
Two-Factor Interactions
DIVISION*PDZIPSAME
DIVISION*DIG_1
PDZIPSAME*DIG_1
PDZIPSAME*AGECAT
PDZIPSAME * REPREPAYEE_1
MOVE_1*PHONE_1
PDZIPSAME * AGECAT
Three-Factor Interactions
DIG_1*DIVISION*PDZIPSAME

TABLE III.4

COOPERATION LOGISTIC PROPENSITY MODEL: REPRESENTATIVE BENEFICIARY SAMPLE

Factors in the Cooperation Model
Main Effects
MOVE_2
SEX
SSI_SSDI_2
DIG
REPREPAYEE_2
PDZIPSAME _2
METRO
DIVISION_2
HISPANICITY
RACE_2
AGECAT_2
TOC_2
PHONE_2
Two-Factor Interactions
RACE_2 * METRO
RACE_2 * SSI_SSDI_2
RACE_2 * DIG
RACE_2 * MOVE_2
RACE_2 * GENDER
RACE_2 * DIVISION_2
DIG * METRO
DIG * MOVE
DIG * SSI_SSDI_2
DIG * REPREPAYEE_2
DIG * SEX
MOVE_2 * SEX
MOVE_2 * METRO
MOVE_2 * DIVISION_2
MOVE_2 * REPREPAYEE_2
MOVE_2 * PHONE_2
MOVE_2 * SSI_SSDI_2
METRO * SEX
METRO * PHONE_2
METRO * TOC_2
Three-Factor Interactions
RACE_2 * MOVE_2 * SEX
RACE_2 * METRO * SEX
RACE_2 * METRO * MOVE_2
RACE_2 * MOVE_2 * SSI_SSDI_2
DIG * MOVE_2 * SEX
DIG * METRO * MOVE_2
MOVE_2 * METRO * SEX

The primary factors are identified by the base variable, followed by the suffix "_1" if levels were collapsed in the variable as it is employed in the location model. If no collapsing was necessary, then no suffix is given. The factors with levels used in the location model include:

1. ***MOVE_1.*** The number of address changes in the past five years; two levels: (1) at least one move, and (2) did not move, or information older than five years, or no information.

2. ***DIG_1.*** Disability diagnostic classification; three levels: (1) mental disability, (2) physical disability (excluding deaf cases), and (3) deaf or unknown.

3. ***REPREPAYEE_1.*** The identity of the payee with respect to the beneficiary; two levels: (1) the beneficiary received benefit payments himself or herself, or from a family member, and (2) an institution received payments on behalf of the beneficiary, or identity of payee not known.

4. ***PDZIPSAME.*** Whether the beneficiary and the applicant for benefits lived in the same zip code; three levels: (1) beneficiary and applicant lived in the same zip code, (2) beneficiary and applicant lived in different zip codes, and (3) information unknown.

5. ***METRO_1.*** Urbanicity of beneficiary's place of residence; three levels: (1) beneficiary lived in metropolitan area, (2) beneficiary lived in nonmetropolitan area adjacent to a metropolitan area of 1 million or more, and (3) beneficiary lived in nonmetropolitan area not adjacent to a metropolitan area with population 1 million or more.

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

7. ***INSTIT.*** Whether beneficiary is institutionalized; two levels: beneficiary is institutionalized, and (2) beneficiary is not institutionalized, or information unknown.

8. ***DIVISION.*** Geographic region (based on U.S. Census divisions) of beneficiary's place of residence; nine levels: (1) Pacific, (2) Mountain, (3) East North Central, (4) West North Central, (5) East South Central, (6) West South Central, (7) South Atlantic, (8) Middle Atlantic, and (9) New England.

9. ***SSIADDP.*** The beneficiary was located at the address of payee obtained from SSI database; two levels: (1) yes, and (2) no.

10. ***LIVING_1.*** Beneficiary's living situation; two levels: (1) beneficiary lives with his or her parents, and (2) beneficiary does not live with his or her parents, or information unknown.

11. ***PHONE_1.*** Number of phone numbers for a beneficiary in the SSA database over past five years; two levels: (1) only one phone number on file, and (2) one or more changes in phone number on SSA file, or information unknown.

12. ***AGECAT.*** Beneficiary's age category; four levels: (1) age in range 18 to 29 years, (2) age in range 30 to 39 years, (3) age in range 40 to 49 years, and (4) age in range 50 to 64 years.

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.3. An expanded form of Table III.3, showing the specific levels of the interactions shown in Table III.3, along with parameter estimates and their standard errors, is provided in Appendix D.

For the cooperation models, the primary factors include:[29]

1. ***MOVE_2.*** The number of address changes in the past five years; two levels: (1) at most one move, and (2) two or more moves, or information older than five years, or no information.
2. ***GENDER (SEX).*** Same as location model definition
3. ***SSI_SSDI_2.*** Beneficiary recipient benefit type; two levels: (1) SSDI only, and (2) SSI only, or both SSI and SSDI.
4. ***DIG.*** Disability diagnostic classification; four levels: (1) mental disability, (2) physical disability (excluding deaf cases), (3) deaf, and (4) unknown.
5. ***REPREPAYEE_2.*** The identity of the payee with respect to the beneficiary; two levels: (1) Family member received benefits on behalf of beneficiary, and (2) the beneficiary received benefit payments himself or herself, or an institution received benefits on behalf of the beneficiary, or unknown.
6. ***PDZIPSAME_2.*** 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.
7. ***METRO.*** Urbanicity of beneficiary's place of residence; six levels: (1) beneficiary lived in metropolitan area with population of 1 million or more, (2) beneficiary lived in metropolitan area with population between 250,000 and 1 million, (3) beneficiary lived in metropolitan area with population less than 250,000, (4) beneficiary lived in nonmetropolitan area adjacent to a metropolitan area of 1 million or more, (5) beneficiary lived in nonmetropolitan area adjacent to a metropolitan area of less

[29] Primary factors that are based on the same base variable as those given in the location model, but with different collapsing of categories, are given the same name except they are followed by an "_2".

than 1 million, and (6) beneficiary lived in nonmetropolitan area not adjacent to any metropolitan area.

8. ***DIVISION_2.*** Geographic region (based on U.S. Census divisions) of beneficiary's place of residence; four levels: (1) South Atlantic, (2) East North Central, (3) West South Central, and (4) other regions of the country that are not South Atlantic, East North Central, or West South Central.
9. ***HISPANICITY.*** Whether the beneficiary was Hispanic or not; two levels: (1) Hispanic, and (2) not Hispanic, or unknown.
10. ***RACE_2.*** Race of the beneficiary; three levels: (1) white, (2) Asian or Pacific islander, and (3) race known to be neither white nor Asian/Pacific Islander, or unknown.
11. ***AGECAT_2.*** Geneficiary's age category; two levels: (1) age in range 40 to 49 years, and (2) age in range 18 to 39 years, or 50 to 64 years.
12. ***TOC_2.*** Beneficiary's type of claim; two levels: (1) disability claim, and (2) survivor claim, or unknown.
13. ***PHONE_2.*** Number of phone numbers on SSA file over past five years; two levels: (1) one or fewer phone changes on SSA file, or unknown, and (2) two or more changes in phone number on SSA file.

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.4. An expanded form of Table III.4, with the specific levels of the interactions shown in Table III.4, along with parameter estimates and their standard errors, is provided in Appendix D.

After adjustments were applied to the sampling weights, the distribution of weights was reviewed to determine if trimming of the sampling weights was necessary. The maximum design effect due to unequal weighting was 1.05, observed with the youngest age group stratum, which indicated that trimming of the weights was unnecessary.

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 the Social Security Administration (SSA).[30] 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 Ticket-to-Work participants, a subset of all SSI/SSDI beneficiaries, which was partitioned based on different payment types in the Ticket-To-Work payment system (traditional vocational rehabilitation, milestone-outcomes, and outcome-only). Ticket participants using the traditional payment system accounted for 81 percent (17,081 of 21,196) of Phase 1 participants and 88 percent (30,254 of 34,312) of Phase 2 participants at the time when the sampling frame was developed. Participants using the milestone-outcomes payment system totaled 3,208 Phase 1 participants (15 percent of all Phase 1 participants) and 3,084 Phase 2 participants (9 percent of all Phase 2 participants). Phase 1 participants using the outcome-only payment system totaled only 907 Phase 1 participants (4 percent of all Phase 1 participants) and 974 Phase 2 participants (3 percent of all Phase 2 participants). As was also noted earlier, the PSUs in the initial sampling design did not contain a sufficient number of participants in the milestone-outcomes or outcome-only payment types for either phase to support analysis tasks. As a result, the clustered sample, consisting of respondents selected within the initial sample design, was supplemented by a sample randomly selected from the entire population of milestone-outcomes and outcome-only participants (this was called the unclustered sample).

[30] 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 ncome (SSDI) beneficiaries.

The clustered sample was part of the original sample design, so all the respondents in the clustered sample 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. In most cases, the respondents who were selected for the in-PSU stratum of the unclustered sample were also in the clustered 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 the next subsection. In addition, respondents who could not be located in the central office[31] 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 in 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 was 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." Sample members with no field follow-up (in the unclustered sample) were not "selected" for field follow-up. This process is analogous to the accepted practice of subsampling on nonrespondents for more intensive effort—in this case, we subsampled cases in the clustered sample for field follow-up. Ineligible-for-field-follow-up cases in the unclustered sample were treated differently than other ineligible cases, regardless of whether the observation was

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

duplicated with a clustered observation. The procedure used to create composite weights (described in the next subsection) was not applied to these cases. Rather, such a case in the unclustered sample would have its weight zeroed out. If such a case was duplicated with one in the clustered sample, the clustered sample case kept its original weight, appropriately adjusted so that the sum of weights was kept the same. The final sample sizes for the participants sample are in Table III.5.

TABLE III.5

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

Sampling Strata (Payment System)	Survey Population[a]	Initial Sample Size	Released Sample
Total Phase 1	34,312	3,528[b]	2,939
1. Traditional payment type	30,254	991	882
2. Milestone-outcomes		1,420	1,084
Clustered sample	3,084	489	438
Unclustered sample	3,084	931	646
In PSUs	824	462	386
Not in PSUs	2,260	469	260
3. Outcome-only			973
Clustered sample	974	168	168
Unclustered sample	974	949	805
In PSUs	168	168	121
Not in PSUs	806	781	973
Total Phase 2	21,196		1,350
1. Traditional Payment Type	17,081	666	437
2. Milestone-Outcomes	3,208	668	436
Clustered sample	1,250	273	216
Unclustered sample	1,958	395	220
In PSUs	154	34	19
Not in PSUs	1,805	361	201
3. Outcome-Only			
Clustered sample	907	86	86
Unclustered sample	907	579	391
In PSUs	86	55	44
Not in PSUs	821	524	347

Source: Sample allocation and counts computed by MPR.

[a] This column reflects weighted totals before compositing.

[b] The initial and final sample sizes include participants using the outcome-only and milestone-outcomes payment systems for which the number obtained from the original sample design was insufficient for analysis. A paired sample design was employed, whereby the participants who were in the PSUs could potentially be selected for both samples.

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) or the Cook County (Chicago) PSU (two SSUs), which were used for the Representative Beneficiary Sample.

1. Initial Weight

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 milestone-outcomes and outcome-only participants, we computed the initial weights based on the selection probability within the two sampling strata (in one of the PSUs or 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. Dual Frame Estimation

In order to obtain estimates for the milestone-outcomes and 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 follow-up. The paired sample design is the methodology used to combine the samples while accounting for these different rules of field follow-up. This requires the creation of composite weights that can be applied to the combined samples.

a. 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, since the clustered and unclustered samples in the Ticket Participant Sample represent the same target population among the Ticket Participants. Separate estimates can be computed from each sample, within each payment type, and combined using the equation

$$(1) \quad Est(Y) = \lambda Y(clustered) + (1 - \lambda)\, Y(unclustered)$$

where *Y(clustered)* is the survey estimate from the clustered sample for the given payment type, *Y(unclustered)* is the survey estimate from the unclustered sample for the given payment type, and λ is an arbitrary constant between 0 and 1. For example, for the Phase 1 milestone-outcomes payment type in the round 2 data, there were 438 in the clustered sample and 646 in the unclustered sample. The estimates to be combined are the proportion of the 438 in the clustered sample who are currently working and the proportion of the 646 in the unclustered sample who are currently working. In practice, of course, it is more complicated than this, because we have to account for the different rules used in the two samples for following up with nonrespondents or unlocated sample members, as will be discussed later. For the sampling variance, *V(Y)*, the estimate is computed using the equation

$$(2) \quad V(Y) = \lambda^2\, V(Y(clustered)) + (1 - \lambda\, V(Y(unclustered))$$

where *V(Y(clustered))* is the sampling variance for the estimate from the clustered sample, and *V(Y(unclustered))* is the sampling variance for the estimate from the unclustered 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:

$$(3)\quad \lambda = 1/V(Y(clustered) / [1 / V(Y(clustered)) + 1/ V(Y(unclustered)]$$

$$= V(Y(unclustered)) / [V(Y(clustered)) + V(Y(unclustered))]$$

In this case, the minimum variance is:

$$(4)\quad V(Y) = [V(Y(clustered)) * V(Y(unclustered))] / [V(Y(clustered)) + V(Y(unclustered))]$$

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.

For this survey round, we used an alternative approach, which is to identify a single lambda that was calculated using sample sizes and design effects due to unequal weighting for the two samples. In particular, λ acts as a weighting factor, with more weight given to the larger sample, with the sample sizes adjusted by the design effect due to unequal weighting. The formula for λ is given by:

$$(5)\quad \lambda = \frac{n(clustered)/deff(clustered)}{n(clustered)/deff(clustered)+n(unclustered)/deff(unclustered)}$$

where *n(clustered)* and *n(unclustered)* are the sample sizes of the clustered and unclustered central office-located samples respectively, and *deff(clustered)* and *deff(unclustered)* are the

design effects due to unequal weighting for the clustered and unclustered central office-located samples, respectively.

b. Application of Composite Weights to Ticket Participant Sample

The population of participants in the relevant payment type 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, those who 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, those who cannot be located by central office locating efforts), only the clustered sample can provide unbiased estimates for this subpopulation, since unclustered sample cases were not eligible for field follow-up.

For the subpopulation that can be located by central office locating efforts, the clustered and unclustered samples can be combined using the compositing method (called a "dual frame" estimation procedure). To compute the composite weight for each sample member in the clustered central office-located sample:

(6) $WT = \lambda\ WT(\text{unclustered central office-located sample weight})$

For units in the unclustered central office-located sample:

(7) $WT = (1 - \lambda)\ WT(\text{clustered central office-located sample weight})$

Conversely, for the subpopulation of persons who could not be found by central office locating efforts, only the clustered sample can be used. In this case, no combining is required, and the clustered weight is used directly:

(8) *WT = 1 * WT(clustered field-located sample weight)*

The sum of weights among cases that were field-located in the clustered sample was adjusted so that the total sum matched the original total sum. Because the weights for each subpopulation sum to the total number of individuals in each subpopulation, the two subpopulations can simply be combined to form the entire target population.

Because of the paucity of sample members in the PSUs in some cases, it was not uncommon for the unclustered sample to be much larger than the clustered sample. When combining samples and creating composite weights, this sometimes resulted in weights with unacceptably high levels of variation. This made trimming necessary to reduce this variation, which is described in a later section.

3. Nonresponse Adjustment

As with the Representative Beneficiary Survey, the sampling weights were adjusted for the sample members who could not be located or who, once located, refused to respond, using weighted logistic regression propensity models. For the milestone-outcomes and outcome-only payment types, the nonresponse adjustments were applied to the composite weights for the clustered and unclustered samples. Roughly equal sample sizes with vastly different population sizes for the three payment types resulted in substantial differences in the magnitude of the weights. It was therefore necessary to fit separate logistic regression models for each payment type and each phase, first for the location adjustment and subsequently for the cooperation adjustment. This resulted in a total of 12 logistic regression models. These models were fitted in the same way as the adjustment models for the Representative Beneficiary Sample, as described in Section B.2 of this chapter. The main factors or attributes affecting our ability to locate and interview Ticket Participant sample members were the same as those used to locate and

interview Representative Beneficiaries, where the specific covariates for each of the 12 logistic models varied as described in subsequent sections.

a. Coding of Survey Dispositions

The scheme used to code respondents included the four general categories described in Section B.2: eligible respondents; ineligible respondents; located nonrespondents, and unlocated sample members.[32]

b. Response Rates

The response rate for the Ticket Participant Sample is 80.4 percent, which is the weighted overall completion rate. This rate is a combination of the Phase 1 weighted overall completion rate (80.9 percent) and the Phase 2 weighted overall completion rate (79.5 percent). It is also the product of the weighted location rate and the weighted completion rate among located sample members. The weighted location rate is 94.5 percent, the combination of the Phase 1 location rate (95.7 percent) and the Phase 2 location rate (92.6 percent). The weighted cooperation rate (the weighted completion rate among located sample members), is 85.0 percent, the combination of the Phase 1 weighted completion rate (84.6 percent) and the Phase 2 weighted completion rate (85.8 percent).

Analogous to the beneficiary sample, the weighted rates are used because the sampling weights vary substantially across the sampling strata, and the weighted rates better reflect the potential for nonresponse bias.

[32] Disposition codes 420 (institutionalized) and 430 (unavailable during field period) were classified as nonrespondent codes in round 2, even though they were considered ineligible codes in round 1. This affected 4 cases in the round 2 participant sample. As a result, the nonresponse adjusted weight for these 4 cases was 0 in round 2, even though a similar response in round 1 would have resulted in a positive weight. Because of the small numbers, the effect on response rates is very small.

c. Factors Related to Location and Response

Tables III.6 – III.11 provide information for selected factors associated with locating a sample member within each phase-payment type combination, and factors associated with response among located sample members. The tables include 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 were determined to be ineligible. The tables also include the weighted location rate, the weighted completion rate among 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.

TABLE III.6

WEIGHTED RESPONSE RATE FOR TICKET PARTICIPANTS SAMPLE, PHASE I STATES, MILESTONE-OUTCOME, BY SELECTED CHARACTERISTICS

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
All	980[a]	932	94.3	761	82.7	78.1
SSI Only, SSDI Only, or Both SSI and SSDI						
SSI only	312	293	92.6	248	86.0	80.2
SSDI only	433	415	96.3	330	80.9	77.7
Both SSI and SSDI	205	196	93.1	167	85.0	79.5
Unknown	30	28	87.0	16	55.5	48.3
SSI or SSDI						
SSI only or in both SSI & SSDI programs	517	489	92.8	415	85.6	79.9
SSDI only or in both SSI & SSDI programs	638	611	95.4	497	82.1	78.2
Constructed Disability Status						
Deaf	16	15	88.1	10	55.8	47.8
Mental	520	494	93.3	423	84.7	79.0
Physical	394	376	96.2	297	82.9	80.3
Unknown	50	47	90.7	31	65.7	60.0
Beneficiary's Age (Four Categories)						
18-29 years	170	160	93.8	131	83.8	78.9
30-39 years	216	205	93.1	177	87.7	81.7
40-49 years	309	293	95.6	228	77.6	74.4
50-64 years	285	274	94.1	225	84.0	79.1

TABLE III.6 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Sex						
Male	496	470	94.3	385	81.2	76.7
Female	484	462	94.3	376	84.1	79.5
Hispanicity						
Hispanic	83	75	81.3	60	80.9	64.7
NonHispanic/unknown	897	857	95.0	701	82.7	78.8
Race						
White	447	429	96.4	344	82.0	79.2
Black	348	331	92.2	285	85.6	79.4
Unknown/other	185	172	92.0	132	79.8	73.1
Living Situation						
Living alone	498	473	92.9	401	85.4	79.8
Living with others/unknown	482	459	95.6	360	80.1	76.6
Did the Applicant for Benefits Live In Same Zip Code as Beneficiary?						
No	97	87	88.1	70	83.6	75.0
Yes	477	456	94.6	383	85.8	81.2
No information	406	389	95.8	308	78.3	75.12
Identity of the Payee with Respect to the Beneficiary						
Beneficiary received beneficiary payments himself or herself	783	747	94.8	597	81.2	77.1
Payee is a family member	156	147	92.6	129	89.2	82.7
Payee is an institution	27	27	100.0	25	84.8	86.4
Other	14	11	82.1	10	90.1	73.4
Changes in Telephone Number						
No changes in last five years	730	695	94.3	568	82.6	78.0
One or more changes in last five years	55	52	92.8	43	77.2	71.8
No information/other	195	185	94.9	150	84.6	80.7
Number of Moves in Last Five Years						
No moves in last five years	307	295	94.4	258	89.4	84.8
One or more moves in last five years	45	41	91.2	31	79.0	71.6
No information/other	628	596	94.6	472	79.7	75.5
Type of Claim						
Survivor	33	33	100.0	24	81.2	81.3
Disabled	647	618	95.0	502	81.4	77.3
Unknown	300	281	92.3	235	85.7	79.7
Address of Payee Obtained from SSI File						
Yes	840	795	93.6	658	83.8	78.6
No	140	137	98.8	103	75.6	75.0
Census Region						
Midwest	157	153	95.3	129	82.9	79.1
Northeast	292	279	94.5	223	78.7	74.2
South	344	326	94.6	272	85.5	81.3
West	187	174	91.9	137	79.4	73.0

TABLE III.6 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Census Division						
East North Central	149	145	94.7	124	85.3	80.8
Middle Atlantic	175	169	95.9	137	79.3	76.0
Mountain	162	152	92.5	117	78.5	72.8
New England	117	110	92.0	86	77.6	71.1
Pacific	25	22	89.1	20	83.6	74.1
South Atlantic	248	231	91.0	191	84.1	77.1
West North Central	8	8	100.0	5	62.7	63.7
West South Central	96	95	99.1	81	87.2	86.5
Metropolitan						
Metropolitan areas of 1 million population or more	797	752	92.0	611	80.0	73.9
Metropolitan areas of 250,000 to 999,999 population	92	90	96.7	75	86.9	83.9
Metropolitan areas of less than 250,000 population	37	37	100.0	29	82.8	82.9
Nonmetropolitan areas adjacent to large metropolitan areas	7	7	100.0	6	85.6	85.7
Nonmetropolitan areas adjacent to medium or small metropolitan areas	27	26	97.5	22	90.2	87.6
Nonmetropolitan areas not adjacent to metropolitan areas	20	20	100.0	18	90.0	90.0
Longitudinal						
Yes	292	272	91.4	222	83.3	76.3
No	688	660	95.3	539	82.4	78.7

Source: NBS, round 2.

[a] This includes both the unclustered and clustered sample cases, some of which are duplicates. It excludes 104 unclustered cases, out of the total of 1,084, that were ineligible for field follow-up.

TABLE III.7

WEIGHTED RESPONSE RATE FOR TICKET PARTICIPANTS SAMPLE, PHASE 1 STATES, OUTCOME-ONLY, BY SELECTED CHARACTERISTICS

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
All	871[a]	846	95.5	668	79.9	75.9
SSI Only, SSDI Only, or Both SSI and SSDI						
SSI only	128	123	92.8	98	79.0	72.1
SSDI only	529	513	96.0	407	81.2	77.7
Both SSI and SSDI	127	124	94.3	99	79.0	74.0
Unknown	87	86	98.2	64	73.6	72.1
SSI or SSDI						
SSI only or in both SSI & SSDI programs	255	247	93.6	197	79.0	73.2
SSDI only or in both SSI & SSDI programs	656	637	95.6	506	80.7	77.0
Constructed Disability Status						
Deaf	20	19	95.3	10	66.1	62.2
Mental	423	415	97.0	332	79.8	77.2
Physical	319	305	92.9	245	82.1	75.9
Unknown	109	107	97.6	81	75.7	73.3
Beneficiary's Age (Four Categories)						
18-29 years	68	67	98.6	54	79.7	77.5
30-39 years	224	216	92.6	160	72.0	66.7
40-49 years	306	296	96.4	238	82.4	79.2
50-64 years	273	267	95.9	216	83.0	79.1
Sex						
Male	457	440	95.4	352	81.4	77.3
Female	414	406	95.6	316	78.1	74.2
Hispanicity						
Hispanic	21	20	94.8	12	73.5	67.0
NonHispanic/unknown	850	826	95.5	656	80.0	76.1
Race						
White	557	544	96.0	436	80.4	76.8
Black	113	111	97.9	86	79.7	77.4
Other/unknown	201	191	92.8	146	78.5	72.6
Living Situation						
Living alone	247	239	93.3	194	80.1	74.0
Living with others/unknown	624	607	96.3	474	79.8	76.6
Did the Applicant for Benefits Live In Same Zip Code as Beneficiary?						
No	68	66	96.6	53	79.8	76.9
Yes	353	341	94.5	276	80.6	75.7
No information	450	439	96.1	339	79.3	76.0

TABLE III.7 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Identity of the Payee with Respect to the Beneficiary						
Beneficiary received beneficiary payments himself or herself	756	735	95.3	581	80.4	76.2
Payee is a family member	73	72	98.5	55	74.0	72.6
Payee is an institution	27	26	96.4	21	79.8	77.0
Other	15	13	89.0	11	82.7	73.3
Changes in Telephone Number						
No changes in last five years	699	677	94.7	536	79.7	75.1
One or more changes in last five years	33	33	100.0	26	76.3	76.1
No information/other	139	136	98.2	106	81.7	79.8
Number of Moves in Last Five Years						
No moves in last five years	239	227	91.9	185	80.4	73.4
One or more moves in last five years	23	23	100.0	17	72.8	72.8
No information/other	609	596	96.8	466	79.9	77.1
Type of Claim						
Survivor	30	29	96.7	22	76.0	73.4
Disabled	675	656	95.7	520	80.4	76.7
Unknown	166	161	94.3	126	78.0	72.9
Address of Payee Obtained from SSI File						
Yes	643	623	95.4	499	80.3	76.3
No	178	173	94.8	131	79.1	74.7
Unknown	50	50	100.0	38	76.2	76.0
Census Region						
Midwest	128	122	94.5	96	74.2	70.2
Northeast	376	372	98.9	302	82.7	81.8
South	216	210	97.0	149	71.6	69.5
West	151	142	85.5	121	88.5	74.8
Census Division						
East North Central	107	102	94.2	85	80.6	76.2
Middle Atlantic	88	86	97.6	66	81.9	80.0
Mountain	79	75	93.0	59	84.4	77.1
New England	288	286	99.3	236	83.0	82.3
Pacific	72	67	77.7	62	93.1	72.4
South Atlantic	205	199	96.8	140	71.1	68.9
West North Central	21	20	95.5	11	50.9	48.6
West South Central	11	11	100.0	9	82.3	81.9

TABLE III.7 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Metropolitan						
Metropolitan areas of 1 million population or more	433	414	92.1	328	82.5	75.3
Metropolitan areas of 250,000 to 999,999 population	105	104	99.1	75	68.8	68.7
Metropolitan areas of less than 250,000 population	88	86	97.8	67	76.4	74.8
Nonmetropolitan areas adjacent to large metropolitan areas	9	8	82.6	5	61.3	51.5
Nonmetropolitan areas adjacent to medium or small metropolitan areas	112	111	99.1	90	81.1	80.4
Nonmetropolitan areas not adjacent to metropolitan areas	124	123	99.2	103	83.9	83.1
Longitudinal						
Yes	440	427	94.9	331	77.6	72.9
No	431	419	96.0	337	81.9	78.7

Source: NBS, round 2.

[a]This includes both the unclustered and clustered sample cases, some of which are duplicates. It excludes 102 unclustered cases, out of the total of 973, that were ineligible for field follow-up.

TABLE III.8

WEIGHTED RESPONSE RATE FOR TICKET PARTICIPANTS SAMPLE, PHASE 1 STATES, TRADITIONAL, BY SELECTED CHARACTERISTICS

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
All	882	846	95.8	722	85.0	81.4
SSI Only, SSDI Only, or Both SSI and SSDI[a]						
SSI only	283	267	94.3	228	85.5	80.4
SSDI only	384	370	96.4	307	82.2	79.1
Both SSI and SSDI	215	209	97.0	187	89.5	86.8
SSI or SSDI						
SSI only or in both SSI & SSDI programs	495	473	95.4	412	87.1	83.1
SSDI only or in both SSI & SSDI programs	589	570	96.6	487	85.0	82.1
Constructed Disability Status						
Deaf	48	46	96.6	35	76.1	73.3
Mental	463	449	96.9	377	83.7	81.0
Physical	344	326	94.7	289	88.3	83.6
Unknown	27	25	91.6	21	81.8	74.4
Beneficiary's Age (Four Categories)						
18-29 years	235	224	95.4	193	85.0	81.0
30-39 years	185	175	95.2	151	86.1	82.0
40-49 years	259	249	95.6	210	84.1	80.3
50-64 years	203	198	97.3	168	85.2	82.8
Sex						
Male	466	442	94.9	385	87.2	82.7
Female	416	404	97.0	337	82.5	80.0
Longitudinal						
Yes	424	405	95.6	332	82.4	78.7
No	458	441	96.2	390	88.5	85.1
Hispanicity						
Hispanic	47	43	91.8	37	87.9	80.8
NonHispanic/unknown	835	803	96.1	685	84.8	81.4
Race						
White	446	430	96.4	372	85.9	82.7
Black	248	236	94.7	206	87.1	82.4
Other/unknown	188	180	96.1	144	80.1	76.9
Living Situation						
Living alone	460	440	95.5	381	86.6	82.8
Living with others/unknown	422	406	96.2	341	83.2	80.0
Did the Applicant for Benefits Live In Same Zip Code as Beneficiary?						
No	70	66	95.3	57	86.4	82.4
Yes	471	451	95.5	402	89.0	85.0
No information	341	329	96.4	263	79.2	76.3

TABLE III.8 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Identity of the Payee with Respect to the Beneficiary						
Beneficiary received beneficiary payments himself or herself	622	597	95.9	514	85.7	82.2
Payee is a family member	186	181	97.1	150	82.3	79.8
Payee is an institution	65	61	93.5	52	85.2	79.7
Other	9	7	84.6	6	89.0	75.2
Changes in Telephone Number						
No changes in last five years	623	601	96.3	516	85.6	82.4
One or more changes in last five years	35	33	94.4	31	95.0	89.8
No information/other	224	212	94.7	175	81.9	77.5
Number of Moves in Last Five Years						
No moves in last five years	314	302	96.1	276	91.2	87.6
One or more moves in last five years	21	18	85.2	16	91.9	78.0
No information/other	547	526	96.1	430	81.3	78.1
Type of Claim						
Survivor	71	68	95.8	61	88.8	85.0
Disabled	537	520	96.7	442	84.5	81.8
Unknown	274	258	94.1	219	85.0	79.8
Address of Payee Obtained from SSI File						
Yes	796	763	95.8	662	86.7	83.0
No	86	83	96.4	60	69.5	67.0
Census Region						
Midwest	295	281	95.7	242	86.4	82.5
Northeast	233	225	96.3	182	80.8	77.7
South	299	287	95.6	249	85.8	82.0
West	55	53	96.4	49	92.9	89.5
Census Division						
East North Central	280	267	95.7	229	86.2	82.4
Middle Atlantic	211	203	96.0	165	80.9	77.6
Mountain	43	41	95.2	38	92.7	88.2
New England	22	22	100.0	17	78.7	78.5
Pacific	12	12	100.0	11	93.4	93.3
South Atlantic	287	277	96.4	241	86.1	82.9
West North Central	15	14	94.5	13	90.8	86.0
West South Central	12	10	83.3	8	80.4	66.7

TABLE III.8 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Metropolitan						
Metropolitan areas of 1 million population or more	450	431	95.6	359	82.7	79.0
Metropolitan areas of 250,000 to 999,999 population	272	260	95.6	230	88.0	84.1
Metropolitan areas of less than 250,000 population	100	98	98.2	83	84.5	82.9
Nonmetropolitan areas adjacent to large metropolitan areas	1	1	100.0	1	100.0	100.0
Nonmetropolitan areas adjacent to medium or small metropolitan areas	40	37	92.3	32	86.5	79.6
Nonmetropolitan areas not adjacent to metropolitan areas	19	19	100.0	17	89.7	89.5

Source: NBS, round 2.

[a] The variable for SSI only, SSD only, and both SSI and SSDI includes slightly fewer unknown cases than the indicators for SSI or SSDI. Supplemental information was obtained to update the values of the former variable that was not used to update the latter variables.

TABLE III.9

WEIGHTED RESPONSE RATE FOR TICKET PARTICIPANTS SAMPLE, PHASE 2 STATES, MILESTONE-OUTCOME, BY SELECTED CHARACTERISTICS

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
All	417[a]	393	91.6	335	84.2	78.0
SSI Only, SSDI Only, or Both SSI and SSDI[b]						
SSI only	153	141	89.3	121	83.5	74.5
SSDI only	162	156	94.3	136	89.5	83.9
Both SSI and SSDI	101	95	90.4	77	76.6	72.9
Unknown	1	1	100.0	1	100.0	100.0
SSI or SSDI						
SSI only or in both SSI & SSDI programs	254	236	89.7	198	80.5	73.8
SSDI only or in both SSI & SSDI programs	259	247	92.7	209	84.3	79.4
Constructed Disability Status						
Deaf	17	14	89.0	9	77.1	67.7
Mental	242	224	88.2	189	81.1	72.7
Physical	140	138	96.0	122	90.2	86.4
Unknown	18	17	98.5	15	78.4	79.3
Beneficiary's Age (Four Categories)						
18-29 years	125	118	85.1	98	81.3	71.4
30-39 years	96	88	91.1	70	79.2	72.0
40-49 years	109	103	95.4	90	85.6	81.7
50-64 years	87	84	97.0	77	92.1	89.3
Sex						
Male	196	186	92.0	161	86.5	81.3
Female	221	207	91.3	174	82.2	75.0
Hispanicity						
Hispanic	4	4	100.0	4	100.0	100.0
NonHispanic/unknown	413	389	91.6	331	84.1	77.9
Race						
White	203	191	90.8	165	84.4	78.0
Black	148	139	93.8	116	81.6	77.3
Other/unknown	66	63	89.8	54	89.4	79.3
Living Situation						
Living alone	239	222	89.4	184	79.2	72.5
Living with others/unknown	178	171	94.4	151	90.4	84.8
Did the Applicant for Benefits Live In Same Zip Code as Beneficiary?						
No	40	36	80.8	29	72.2	64.4
Yes	220	207	92.9	172	82.1	76.4
No information	157	150	92.7	134	90.8	84.1

TABLE III.9 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Identity of the Payee with Respect to the Beneficiary						
Beneficiary received beneficiary payments himself or herself	287	271	92.3	236	87.0	80.0
Payee is a family member	92	88	95.1	72	83.0	78.9
Payee is an institution	28	24	77.2	18	60.0	54.5
Other	10	10	100.0	9	96.7	96.9
Changes in Telephone Number						
No changes in last five years	280	267	93.0	234	86.1	81.2
One or more changes in last five years	21	19	92.6	15	71.9	68.8
No information/other	116	107	87.8	86	81.6	71.2
Number of Moves in Last Five Years						
No moves in last five years	149	140	90.5	120	80.5	75.0
One or more moves in last five years	19	17	91.4	13	67.3	63.7
No information/other	249	236	92.4	202	88.1	81.1
Type of Claim						
Survivor	44	42	97.6	34	88.9	86.0
Disabled	221	211	92.0	181	83.9	78.8
Unknown	152	140	89.2	120	83.3	74.3
Address of Payee Obtained from SSI File						
Yes	374	351	91.8	296	82.7	76.8
No	43	42	90.3	39	95.5	86.5
Census Region						
Midwest	137	124	85.8	109	84.4	74.1
Northeast	21	21	100.0	19	91.7	90.8
South	245	234	95.3	194	82.8	79.0
West	14	14	100.0	13	91.1	93.0
Census Division						
East North Central	115	105	88.4	93	84.5	76.7
East South Central	44	44	100.0	38	87.5	87.4
Middle Atlantic	10	10	100.0	9	88.6	87.7
Mountain	13	13	100.0	12	90.5	92.5
New England	11	11	100.0	10	95.6	95.2
Pacific	1	1	100.0	1	100.0	100
South Atlantic	30	28	93.3	24	84.8	79.8
West North Central	22	19	74.9	16	83.4	63.0
West South Central	171	162	93.3	132	79.2	73.8

TABLE III.9 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Metropolitan						
Metropolitan areas of 1 million population or more	262	248	94.3	210	82.0	79.2
Metropolitan areas of 250,000 to 999,999 population	46	44	94.5	35	76.3	73.1
Metropolitan areas of less than 250,000 population	47	45	90.3	40	88.8	80.0
Nonmetropolitan areas adjacent to large metropolitan areas	27	24	85.2	22	92.2	78.8
Nonmetropolitan areas adjacent to medium or small metropolitan areas	17	16	94.1	13	83.1	76.5
Nonmetropolitan areas not adjacent to metropolitan areas	18	16	77.0	15	94.6	72.4

Source: NBS, round 2.

[a] This includes both the unclustered and clustered sample cases, some of which are duplicates. It excludes 19 unclustered cases, out of the total of 436, that were ineligible for field follow-up.

[b] The variable for SSI only, SSD only, and both SSI and SSDI includes slightly fewer unknown cases than the indicators for SSI or SSDI. Supplemental information was obtained to update the values of the former variable that was not used to update the latter variables.

TABLE III.10

WEIGHTED RESPONSE RATE FOR TICKET PARTICIPANTS SAMPLE, PHASE 2 STATES, OUTCOME-ONLY, BY SELECTED CHARACTERISTICS

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
All	441[a]	427	96.4	322	76.5	74.2
SSI Only, SSDI Only, or Both SSI and SSDI[b]						
SSI only	66	64	97.3	51	82.5	80.0
SSDI only	310	301	96.4	217	72.6	70.6
Both SSI and SSDI	64	62	96.8	54	87.6	84.5
Unknown	1	0	0	0	0	0
SSI or SSDI						
SSI only or in both SSI & SSDI programs	128	124	97.0	104	85.3	82.5
SSDI only or in both SSI & SSDI programs	358	347	96.3	259	75.4	73.0
Constructed Disability Status						
Deaf	19	17	83.8	9	40.5	38.4
Mental	214	208	96.6	153	76.0	73.3
Physical	185	180	97.6	144	81.5	79.4
Unknown	23	22	95.1	16	71.0	68.8
Beneficiary's Age (Four Categories)						
18-29 years	48	47	97.8	35	71.8	70.6
30-39 years	111	108	97.1	83	78.6	77.2
40-49 years	168	162	95.0	118	74.7	71.3
50-64 years	114	110	97.5	86	79.6	77.5
Sex						
Male	226	217	94.8	158	73.2	70.2
Female	215	210	98.0	164	80.0	78.2
Hispanicity						
Hispanic	10	10	100.0	9	87.1	88.3
NonHispanic/unknown	431	417	96.3	313	76.3	73.9
Race						
White	271	260	95.5	187	73.5	70.7
Black	88	86	97.5	67	80.1	77.8
Unknown	82	81	98.7	68	83.7	82.6
Living Situation						
Living alone	120	116	96.8	96	84.3	81.4
Living with others/unknown	321	311	96.2	226	73.2	71.1
Did the Applicant for Benefits Live In Same Zip Code as Beneficiary?						
No	36	35	99.1	28	79.0	78.1
Yes	161	154	94.7	127	82.4	78.2
No information	244	238	97.1	167	72.5	71.0

TABLE III.10 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Identity of the Payee with Respect to the Beneficiary						
Beneficiary received beneficiary payments himself or herself	369	357	96.3	269	76.4	74.1
Payee is a family member	50	49	97.8	35	71.6	69.4
Payee is an institution	11	11	100.0	10	96.6	96.6
Other	11	10	89.2	8	83.6	74.6
Changes in Telephone Number						
No changes in last five years	364	352	96.1	263	75.3	72.9
One or more changes in last five years	10	9	90.0	8	89.5	80.0
No Information/other	67	66	98.6	51	80.6	79.3
Number of Moves in Last Five Years						
No moves in last five years	110	105	96.5	84	80.1	77.5
One or more moves in last five years	9	8	88.9	8	100.0	88.9
No information/other	322	314	96.6	230	74.7	72.6
Type of Claim						
Survivor	30	30	100.0	22	74.1	73.9
Disabled	347	334	95.6	250	75.5	72.7
Unknown	64	63	98.6	50	82.2	80.8
Address of Payee Obtained from SSI File						
Yes	326	315	96.3	245	78.8	76.2
No	115	112	96.9	77	68.8	67.2
Census Region						
Midwest	83	79	96.3	64	85.3	82.0
Northeast	246	239	95.6	173	71.7	69.3
South	99	97	98.7	75	78.2	77.2
West	13	12	92.3	10	84.0	76.9
Census Division						
East North Central	42	41	99.0	29	69.3	68.0
East South Central	39	38	97.5	28	74.9	73.0
Middle Atlantic	58	53	88.1	37	68.6	61.0
Mountain	12	11	91.7	9	82.7	75.0
New England	188	186	98.2	136	72.7	72.2
Pacific	1	1	100.0	1	100.0	100.0
South Atlantic	35	35	100.0	26	74.8	74.7
West North Central	41	38	94.9	35	94.5	89.8
West South Central	25	24	98.8	21	88.2	87.1

TABLE III.10 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Metropolitan						
Metropolitan areas of 1 million population or more	159	151	94.5	114	75.5	71.6
Metropolitan areas of 250,000 to 999,999 population	171	167	96.6	122	72.9	71.4
Metropolitan areas of less than 250,000 population	33	31	93.9	26	84.0	78.8
Nonmetropolitan areas adjacent to large metropolitan Areas	12	12	100.0	9	90.2	89.7
Nonmetropolitan areas adjacent to medium or small metropolitan areas	45	45	100.0	36	80.2	80.0
Nonmetropolitan areas not adjacent to metropolitan areas	21	21	100.0	15	70.7	70.7

Source: NBS, round 2.

[a] This includes both the unclustered and clustered sample cases, some of which are duplicates. It excludes 36 unclustered cases, out of the total of 477, that were ineligible for field follow-up.

[b] The variable for SSI only, SSD only, and both SSI and SSDI includes slightly fewer unknown cases than the indicators for SSI or SSDI. Supplemental information was obtained to update the values of the former variable that was not used to update the latter variables.

TABLE III.11

WEIGHTED RESPONSE RATE FOR TICKET PARTICIPANTS SAMPLE, PHASE 2 STATES, TRADITIONAL, BY SELECTED CHARACTERISTICS

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
All	437	404	92.6	348	86.4	80.0
SSI Only, SSDI Only, or Both SSI and SSDI[a]						
SSI only	114	109	95.9	92	83.5	80.1
SSDI only	216	196	90.9	170	87.2	79.3
Both SSI and SSDI	107	99	92.8	86	87.6	81.4
SSI or SSDI						
SSI only or in both SSI & SSDI programs	219	206	94.3	176	85.4	80.6
SSDI only or in both SSI & SSDI programs	320	293	91.8	254	87.3	80.1
Constructed Disability Status						
Deaf	19	14	77.5	11	81.6	61.9
Mental	237	220	92.8	193	87.5	81.2
Physical	165	155	94.1	131	85.4	80.3
Unknown	16	15	93.3	13	86.2	81.5
Beneficiary's Age (Four Categories)						
18-29 years	123	116	94.5	100	86.7	82.0
30-39 years	96	83	86.9	66	79.7	68.9
40-49 years	123	116	94.2	104	90.0	84.8
50-64 years	95	89	94.1	78	88.0	82.8
Sex						
Male	217	198	91.4	170	86.3	78.9
Female	220	206	93.8	178	86.5	81.2
Hispanicity						
Hispanic	4	3	73.8	2	73.5	54.0
NonHispanic/unknown	433	401	92.8	346	86.5	80.3
Race						
White	270	253	94.0	218	86.5	81.2
Black	94	83	88.4	72	86.4	76.5
Other/unknown	73	68	93.0	58	86.0	80.2
Living Situation						
Living alone	204	193	94.8	164	85.0	80.7
Living with others/unknown	233	211	90.7	184	87.6	79.5
Did the Applicant for Benefits Live In Same Zip Code as Beneficiary?						
No	41	39	95.0	30	78.1	73.9
Yes	222	205	92.7	180	88.0	81.5
No information	174	160	92.0	138	86.4	79.6

TABLE III.11 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Identity of the Payee with Respect to the Beneficiary						
Beneficiary received beneficiary payments himself or herself	306	282	92.4	240	85.6	79.1
Payee is a family member	96	90	93.9	80	89.1	83.5
Payee is an institution	24	23	95.8	22	95.6	91.6
Other	11	9	82.9	6	63.4	52.4
Changes in Telephone Number						
No changes in last five years	316	290	92.0	251	87.2	80.2
One or more changes in last five years	14	13	92.7	10	75.4	70.7
No information/other	107	101	94.6	87	85.5	80.9
Number of Moves in Last Five Years						
No moves in last five years	146	132	90.9	116	88.9	80.6
One or more Moves in last five years	12	11	92.1	10	89.0	83.5
No information/other	279	261	93.6	222	85.0	79.6
Type of Claim						
Survivor	46	42	91.4	38	90.5	83.1
Disabled	281	257	91.7	221	86.7	79.4
Unknown	110	105	95.7	89	83.8	80.3
Address of Payee Obtained from SSI File						
Yes	384	354	92.4	306	86.6	80.1
No	53	50	94.2	42	84.6	79.9
Census Region						
Midwest	216	201	93.2	171	84.8	79.3
Northeast	19	18	94.1	17	95.6	89.8
South	166	152	91.8	133	88.8	81.2
West	36	33	91.7	27	82.0	75.0
Census Division						
East North Central	200	185	92.6	157	84.5	78.6
Middle Atlantic	41	34	85.0	30	89.3	75.4
Mountain	8	7	83.4	6	86.1	71.5
New England	36	33	91.7	27	82.0	75.0
Pacific	11	11	100.0	11	100.0	100.0
South Atlantic	78	72	92.5	66	92.2	85.0
West North Central	16	16	100.0	14	87.0	86.5
West South Central	47	46	97.5	37	80.8	78.8

TABLE III.11 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Metropolitan						
Metropolitan areas of 1 million population or more	195	183	93.5	153	84.1	78.7
Metropolitan areas of 250,000 to 999,999 population	57	53	93.4	45	84.3	79.2
Metropolitan areas of less than 250,000 population	71	60	85.6	52	87.5	74.4
Nonmetropolitan areas adjacent to large metropolitan areas	41	40	97.9	35	86.8	84.8
Nonmetropolitan areas adjacent to Medium or small metropolitan areas	15	14	93.6	14	100.0	93.6
Nonmetropolitan areas not adjacent to metropolitan areas	58	54	93.1	49	90.5	84.5

Source: NBS, round 2.

[a] The variable for SSI only, SSD only, and both SSI and SSDI includes slightly fewer unknown cases than the indicators for SSI or SSDI. Supplemental information was obtained to update the values of the former variable that was not used to update the latter variables.

d. Propensity Models for Weight Adjustments

As with the Representative Beneficiary Sample, the weight adjustments used in the Ticket Participant Sample were based on predicted propensities from a logistic regression 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. Interactions to be considered for inclusion in model development were identified using CHAID, as described in the model-fitting section for the Representative Beneficiary Sample.

After identifying a smaller pool of main effects and interactions for potential inclusion in the final model using backward and forward stepwise regressions, a set of models was statistically

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 the stepwise procedures, we again included variables or interactions that had a statistical significance level (alpha level) of 0.30 or lower (instead of the standard 0.05). Once the candidate list of main effects and interactions was determined, a thorough model-fitting process was used to determine a parsimonious model with few very small propensities. The main effects and the interactions in the models are summarized in Tables III.12 and III.13 for locating a sample member, and in Table III.14 and III.15 for cooperation among located sample members. The R-squared values for the 12 logistic models are given in Table III.16. The unadjusted R-squared value for the location models ranges from a low of 0.015 (0.051 when rescaled to have a maximum of 1) up to 0.156 (0.258 when rescaled to have a maximum of 1). The unadjusted R-squared value for the nonresponse models ranges from a low of 0.059 (0.145 when rescaled as above) up to 0.155 (0.354 when rescaled). These values are similar to those observed for other response propensity modeling efforts using logistic regression with design-based sampling weights. The levels of concordant and discordant pairs, and the p-values for the Hosmer-Lemeshow goodness-of-fit test, are given in Table III.17.

Although the minimum proportion of concordant pairs is 58.7 (Phase 1 traditional location model), and the maximum proportion of discordant pairs is 36.7 (Phase 1 outcome only cooperation model), the difference between the proportion of concordant pairs and the proportion of discordant pairs exceeds 0.3 for all models. The minimum p-value associated with the Hosmer-Lemeshow goodness-of-fit test is 0.184, indicating no evidence of lack of fit for any of the models.

TABLE III.12

LOCATION LOGISTIC PROPENSITY MODELS: PHASE 1 TICKET PARTICIPANT SAMPLES

Factors in the Milestone-Outcome Location Model
Main Effects
MOVE
PDZIPSAME
METRO
REGION
PHONE
AGECAT
SSI_SSDI
RACE
HISPANIC
LONG
Two-Factor Interactions
MOVE*PDZIPSAME
MOVE*HISPANIC
PDZIPSAME*REGION
AGECAT*LONG
RACE*PHONE
Factors in the Outcome-Only Location Model
Main Effects
DIG
METRO
SEX
REGION
PHONE
SSI_SSDI
TOC
Two-Factor Interactions
REGION*SEX
REGION*SSI_SSDI
Factors in the Traditional Location Model
Main Effects
MOVE
DIG
SEX
AGECAT
SSI_SSDI

TABLE III.13

LOCATION LOGISTIC PROPENSITY MODELS: PHASE 2 TICKET PARTICIPANT SAMPLES

Factors in the Milestone-Outcome Location Model
Main Effects
DIG
REPREPAYEE
PDZIPSAME
METRO
SEX
REGION
LIVING
AGECAT
Two-Factor Interactions
PDZIPSAME*REGION
REPREPAYEE*AGECAT
SEX*REGION
Factors in the Outcome-Only Location Model
Main Effects
DIG
METRO
SEX
REGION
SSI_SSDI
TOC
RACE
Two-Factor Interactions
DIG*SEX
Factors in the Traditional Location Model
Main Effects
DIG
METRO
LIVING
AGECAT
SSI_SSDI
RACE

TABLE III.14

COOPERATION LOGISTIC PROPENSITY MODELS: PHASE 1 TICKET PARTICIPANT SAMPLES

Factors in the Milestone-Outcome Cooperation Model
Main Effects
MOVE
DIG
REPREPAYEE
PDZIPSAME
REGION
LIVING
PHONE
AGECAT
RACE
LONG
Two-Factor Interactions
MOVE*DIG
MOVE*LIVING
DIG*PDZIPSAME
DIG*RACE
DIG*LONG
REPREPAYEE*LIVING
REPREPAYEE*PHONE
PDZIPSAME*RACE
REGION*AGECAT
AGECAT*RACE
RACE*LONG
Three-Factor Interactions
DIG*PDZIPSAME*RACE
DIG*PDZIPSAME*LONG
Factors in the Outcome-Only Cooperation Model
Main Effects
REPREPAYEE
METRO
SEX
REGION
LIVING
AGECAT
SSI_SSDI
SSIADDP
LONG
Two-factor interactions
REGION*AGECAT
REGION*LONG

TABLE III.14 *(continued)*

Factors in the Traditional Cooperation Model
Main Effects
MOVE
DIG
REPREPAYEE
PDZIPSAME
METRO
SEX
REGION
LIVING
SSI_SSDI
SSIADDP
TOC
RACE
HISPANIC
Two-Factor Interactions
MOVE*DIG
MOVE*METRO
DIG*PDZIPSAME
DIG*SSI_SSDI
METRO*SSIADDP

TABLE III.15

COOPERATION LOGISTIC PROPENSITY MODELS: PHASE 2 TICKET PARTICIPANT SAMPLES

Factors in the Milestone-Outcome Cooperation Model
Main Effects
DIG
REPREPAYEE
SEX
REGION
LIVING
PHONE
TOC
Two-Factor Interactions
DIG*REPREPAYEE
REGION*LIVING
Factors in the Outcome-Only Cooperation Model
Main Effects
DIG
PDZIPSAME
SEX
REGION
SSI_SSDI
RACE
Two-Factor Interactions
DIG*RACE
PDZIPSAME*REGION
SEX*SSI_SSDI
Factors in the Traditional Cooperation Model
Main Effects
MOVE
DIG
PDZIPSAME
METRO
REGION
AGECAT
TOC
Two-Factor Interactions
DIG*PDZIPSAME
METRO*TOC
REGION*AGECAT

TABLE III.16

UNADJUSTED AND ADJUSTED R-SQUARED VALUES FOR LOGISTIC PROPENSITY MODELS

Model			Unadjusted R-Squared Value	Adjusted R-Squared Value
Phase	Payment Type	Location or Cooperation		
1	Milestone-Outcome	Location	0.084	0.236
1	Milestone-Outcome	Cooperation	0.156	0.258
1	Outcome-Only	Location	0.090	0.292
1	Outcome-Only	Cooperation	0.071	0.112
1	Traditional	Location	0.015	0.051
1	Traditional	Cooperation	0.103	0.180
2	Milestone-Outcome	Location	0.155	0.354
2	Milestone-Outcome	Cooperation	0.110	0.190
2	Outcome-Only	Location	0.069	0.258
2	Outcome-Only	Cooperation	0.092	0.138
2	Traditional	Location	0.059	0.145
2	Traditional	Cooperation	0.087	0.159

TABLE III.17

PROPORTIONS OF CONCORDANT AND DISCORDANT PAIRS AND HOSMER-LEMESHOW P-VALUES FOR LOGISTIC PROPENSITY MODELS

Model					
Phase	Payment Type	Location or Cooperation	Proportion Concordant	Proportion Discordant	Hosmer-Lemeshow p-Value
1	Milestone-Outcome	Location	72.3	25.1	0.630
1	Milestones-Outcome	Cooperation	71.2	28.1	0.734
1	Outcome-Only	Location	76.5	20.9	0.225
1	Outcome-Only	Cooperation	62.3	36.7	0.452
1	Traditional	Location	58.7	27.9	0.403
1	Traditional	Cooperation	74.6	24.9	0.361
2	Milestone-Outcome	Location	79.3	19.5	0.184
2	Milestone-Outcome	Cooperation	65.1	33.7	0.988
2	Outcome-Only	Location	78.5	17.6	0.923
2	Outcome-Only	Cooperation	65.5	31.8	0.681
2	Traditional	Location	73.5	24.2	0.829
2	Traditional	Cooperation	73.3	25.3	0.649

The primary factors in the location models are given below, with potential levels used in the models. Only the base variables with all possible levels are given. Details about how these levels were collapsed for each model are given in Appendix D.

1. ***MOVE.*** The number of address changes in the past five years; possible levels: (1) no moves, (2) one move, (3) two or more moves, and (4) information older than five years or no information.

2. ***DIG.*** Disability diagnostic classification; possible levels: (1) mental disability, (2) physical disability (excluding deaf cases), (3) deaf, and (4) unknown.

3. ***REPREPAYEE.*** The identity of the payee with respect to the beneficiary; possible levels: (1) the beneficiary received payments himself or herself, (2) a family member received benefits on behalf of the beneficiary, and (3) an institution received payments on behalf of the beneficiary, or identity of payee not known

4. ***PDZIPSAME.*** Whether the beneficiary and the applicant for benefits lived in the same zip code; possible levels: (1) beneficiary and applicant lived in the same zip code, (2) beneficiary and applicant lived in different zip codes, and (3) information unknown.

5. ***METRO.*** Urbanicity of beneficiary's place of residence; possible levels: (1) beneficiary lived in metropolitan area of 1 million or more residents, (2) beneficiary lived in metropolitan area of 250,000 to 1 million residents, (3) beneficiary lived in metropolitan area of less than 250,000 residents, (4) beneficiary lived in nonmetropolitan area adjacent to a metropolitan area of 1 million or more, (5) beneficiary lived in nonmetropolitan area adjacent to a metropolitan area of less than 1 million, and (6) beneficiary lived in nonmetropolitan area not adjacent to any metropolitan area.

6. ***GENDER (SEX); possible levels.*** (1) male, and (2) female .

7. ***REGION.*** Geographic region (based on U.S. Census regions) of beneficiary's place of residence; possible levels: (1) West, (2) Midwest, (3) South, and (4) Northeast.

8. ***LIVING.*** Beneficiary's living situation; possible levels: (1) beneficiary lives alone, (2) beneficiary lives with his or her parents, and (3) beneficiary does not live alone or with his or her parents, or information unknown.

9. ***PHONE.*** Number of phone numbers on SSA file over past five years; possible levels: (1) only one phone number on file, (2) one change in phone number on SSA file, (3) two or more changes in phone number on SSA file, and (4) information unknown.

10. ***AGECAT.*** Beneficiary's age category. Possible levels: (1) age in range 18 to 29 years, (2) age in range 30 to 39 years, (3) age in range 40 to 49 years, and (4) age in range 50 to 64 years.

11. ***SSI_SSDI.*** Beneficiary status; possible levels: (1) SSI only, (2) SSDI only, or (3) both SSI and SSDI.
12. ***TOC.*** Type of claim; possible levels: (1) survivor claim, (2) disability claim, and (3) type of claim unknown
13. ***RACE.*** Race; possible levels: (1) white, (2) black, (3) Asian or Pacific islander, and (4) not white, black, or Asian/Pacific islander, or unknown.
14. ***HISPANIC.*** Hispanic; possible levels: (1) Hispanic, and (2) not Hispanic, or unknown.
15. ***LONG.*** Longitudinal case; possible levels: (1) longitudinal case, and (2) not a longitudinal case

Various interactions among these variables were also included in the model for locating the sample member. A list of the main effects using variable names listed above, as well as interactions, is provided in Tables III.12 and III.13. An expanded form of Table III.12, with the specific levels of the main effects for each model and the interactions shown in Tables III.12 and III.13, along with parameter estimates and their standard errors, is provided in Appendix D. The primary factors in the cooperation models are given below. As with the location models, only the base variables are given. Since all the levels for the base variable were given in the discussion of the location models, they are not provided here. Details about how these levels were collapsed for each model are given in Appendix D:

1. ***MOVE.*** The number of address changes in the past five years
2. ***DIG.*** Disability diagnostic classification
3. ***REPREPAYEE.*** The identity of the payee with respect to the beneficiary
4. ***PDZIPSAME.*** Whether the beneficiary and the applicant for benefits lived in the same zip code
5. ***METRO.*** Urbanicity of beneficiary's place of residence
6. ***GENDER (SEX)***
7. ***REGION.*** Geographic region (based on U.S. Census regions) of beneficiary's place of residence
8. ***LIVING.*** Beneficiary's living situation

9. ***PHONE.*** Number of phone numbers on SSA file over past five years
10. ***AGECAT.*** Beneficiary's age category.
11. ***SSI_SSDI.*** Beneficiary status
12. ***SSIADDP.*** Address of payee obtained from SSI file
13. ***TOC.*** Type of claim
14. ***RACE***
15. ***HISPANIC***
16. ***LONG.*** Longitudinal case

Once again, various interactions among these variables were also included in the model for the cooperation of the sample members. A list of the main effects using variable names listed above, as well as interactions, is provided in Tables III.14 and III.15. An expanded form of Tables III.14 and III.15, with levels appropriately collapsed for each model and the specific levels of the interactions shown in Tables III.14 and III.15, along with parameter estimates and their standard errors, is provided in Appendix D.

4. Trimming

After adjustments were applied to the sampling weights, the distribution of weights was reviewed to determine if trimming of the sampling weights was necessary. Because of the wide variation in the magnitude of the weights due to the use of the composite weights in the milestone-outcome and outcome-only payment types, trimming was sometimes necessary to increase precision of survey estimates. However, we minimize the extent of trimming to reduce the potential for bias in the survey estimates. The design effects associated with each of the six phase-payment type combinations before and after trimming, before poststratification, are presented in Table III.18. Design effects were calculated separately within trimming strata, which were in turn defined within phase-payment type combinations. In general, the trimming strata were defined according to whether the observation was in the clustered or unclustered

sample and whether the sample was part of the longitudinal or supplemental sample. For unclustered cases, the trimming strata were further subdivided according to whether the sample case was in a PSU or not, and whether the frame used to select the sample value was the round 1 frame or the round 2 frame. The strata within which trimming was employed are given in Table III.18. If no trimming was employed for a phase-payment type combination, the maximum design effect across all trimming strata is presented. In that instance, the stratum associated with that maximum design effect is not presented, since in most cases, when no trimming is required the design effects do not differ significantly across trimming strata.

TABLE III.18

DESIGN EFFECTS BEFORE AND AFTER TRIMMING, WITHIN TRIMMING STRATA, FOR SIX PHASE-PAYMENT TYPE COMBINATIONS

Payment Type and Phase		Trimming Stratum Where Trimming Occurred	Design Effect	
Phase	Payment Type		Before Trimming	After Trimming
1	Milestone-Outcome	Clustered Sample, Supplemental	3.37	3.13
1	Outcome-Only	Clustered Sample, Longitudinal	5.03	3.72
1	Traditional	No Trimming	1.06 (maximum)	1.06 (maximum)
2	Milestone- Outcome	No Trimming	1.89 (maximum)	1.89 (maximum)
2	Outcome-Only	Clustered Sample	3.40	2.86
2	Traditional	No Trimming	1.04 (maximum)	1.04 (maximum)

5. Post-Stratification

After the nonresponse adjustment and trimming, the weights were poststratified to the population age and gender totals for each payment type obtained from the SSA sampling frame. This sampling frame included all SSI or SSDI beneficiaries for each payment type within the Ticket Participant population. The distributions of weights within each phase and payment type combination were rechecked to determine if more weight trimming was necessary. No additional weight trimming was required.

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 either 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.

If any missing values existed in variables used to define imputation classes, two different strategies were employed: (1) match recipients to donors who were also missing the value for the covariate; or (2) employ separate hot decks depending upon the availability of the variables defining the imputation classes. In the first instance, the level defined as the missing value was

treated as a separate level. In other words, if a recipient was missing a value for a variable defining an imputation class, then the donor also was missing the value for that variable. This strategy was employed if there were large numbers of donors and recipients missing the covariate in question. In the second instance, for a given recipient, a variable was only used to define the imputation class for that recipient if there was no missing value for that variable. The variables used to define an imputation class for each recipient would depend upon what values were nonmissing among those variables.

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 imputed variable. 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 nine 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; (6) constructed from other variables with imputed values; (7) round 1 data. In most cases, the logical assignments were done using imputed values.[33] 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

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

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 race and ethnicity of the respondents. 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.

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 with imputation classes defined by the zip code of each sample member.

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, 6 imputations from SSA's administrative data, 16 longitudinal imputations, 94 imputations from random hot deck	117	8,106	1.44
C_Race_i	Race	23 longitudinal imputations, 141 imputations from SSA's administrative data, 150 imputations from random hot deck	314	8,106	3.87

Source: NBS, round 2.

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, with a separate category for respondents reporting multiple races. 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 with imputation classes defined by the zip code of each sample member. If the respondent was a longitudinal case, then the imputed value from round 1 was used.

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 of whether the respondent was currently working. These imputed variables include three variables that 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. Additional disability status variables include 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. As with race and ethnicity, the age when the disability was first diagnosed cannot change from one round to the next. For 14 missing values among longitudinal cases, this age variable was obtained from round 1 data. 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 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	104 sequential hot deck	104[a]	7,388	1.40
C_MainConColDiagGrp_i	Main condition diagnosis group collapsed	104 sequential hot deck	104	7,388	1.40
C_MainConBodyGroup_i	Main condition body group	104 sequential hot deck	104	7,388	1.40
C_Disage_i	Age at onset of disability	208 sequential hot deck; 21 from longitudinal data	229	8,106	2.83
C_Adultchild_onset_i	Adult/child onset of disability	29 logical	29	8,106	0.36
B24_i	Currently working	8 random hot deck	8	8,106	0.09

Source: NBS, round 2.

[a] Imputations include 31 cases coded as don’t know or refused on B1 (condition exists which limit respondent's ability to work). To match the procedure used in round 1, these cases were all assumed to have a value of 1, indicating such a condition existed. The remaining 73 cases were code as don't know, refused, condition not reported, or uncodeable for one or more of the constituent diagnosis 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.

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 these were combined to form the variables in the table. 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."

TABLE IV.3

CURRENT JOBS IMPUTATIONS

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
C1_i	Number of current jobs	5 random hot deck	5	1,769	0.28
C2_1_1d_i	Main current job SOC code to one digit	12 random hot deck[a]	12	1,769	0.68
C8_1_i	Hours per week usually worked at current main job	38 random hot deck[b]; 5 logical	43	1,769	2.43
C_TotCurWkHrs_i	Total weekly hours at all current jobs	39 random hot deck[c], 8 constructed from imputed variables	47[d]	1,769	2.66
C_TotCurHrMnth_i	Total hours per month at all current jobs	47 constructed from imputed variables	47	1,769	2.66
C_MainCurJobHrPay_i	Hourly pay at current main job	12 logical, 209 constructed from imputed variables	221	1,769	12.49
C_MainCurJobMnthPay_i	Monthly pay at current main job	14 logical, 12 imputed by distributional assumptions, 197 constructed from imputed variables	223	1,769	12.61
C_TotCurMnthPay_i	Total monthly salary all current jobs	22 logical, 197 sequential hot deck, 15 constructed from imputed variables	234	1,769	13.23

Source: NBS, round 2.

[a] Imputations for current job variables include 5 cases coded as don't know or refused in B24, which were imputed as currently working in B24_i.

[b] If C8_1_i was imputed by hot deck and the respondent had only one job, then the flag indicated that C_TotCurWkHrs_i was imputed by hot deck, even though this variable was not processed in the hot deck program.

[c] The 5 "logically assigned" values are cases with 2 or more jobs, where one or more of the variables associated with the second, third, or fourth jobs may or may not be nonmissing. The values were assigned medians of similar respondents who were missing or not missing these three variables in the same way.

[d] The 46 missing values do not include four cases 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 these cases 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)

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 imputed gross monthly pay for 22 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 where imputations were applied are in Section I of the National Beneficiary Survey questionnaire. 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	16 hot deck	16	8,106	0.19
I9_i	Current health	48 hot deck	48	8,106	0.59
I17a_i	Wear glasses	20 hot deck	20	8,106	0.25
I17b_i	Difficulty seeing with glasses	17 logical, 36 hot deck	53	5,219	1.02
I18_i	Difficulty seeing no glasses	46 logical, 68 hot deck	114	2,887	3.95
I19_i	Uses special equipment because of difficulty seeing	78 logical, 10 hot deck	88	3,443	2.56
I21_i	Difficulty hearing	1 logical, 38 hot deck	39	8,106	0.48
I22_i	Able to hear normal conversation	34 logical, 25 hot deck	59	1,507	3.92
I23_i	Uses special equipment because of difficulty hearing	34 logical, 5 hot deck	39	1,507	2.59
I25_i	Difficulty having speech understood	6 logical, 47 hot deck	53	8,106	0.65
I26_i	Able to have speech understood at all	31 logical, 19 hot deck	50	2,279	2.19
I27_i	Uses special equipment because of difficulty speaking	31 logical, 9 hot deck	40	2,279	2.19
I29_i	Difficulty walking without assistance	12 logical, 47 hot deck	59	8,106	0.73
I30_i	Able to walk ¼ mile	22 logical, 75 hot deck	97	3,531	2.75
I31_i	Uses special equipment because of difficulty walking	22 logical, 16 hot deck	38	3,531	1.08
I33_i	Difficulty climbing 10 steps	9 logical, 76 hot deck	85	8,106	1.05
I34_i	Able to climb 10 steps at all	43 logical, 47 hot deck	90	3,664	2.46
I35_i	Difficulty lifting and carrying 10 lbs.	6 logical, 52 hot deck	58	8,106	0.72

TABLE IV.4 *(continued)*

Variable Name	Description	Imputation Method[a]	Number Missing	Number Eligible	Percent Imputed
I36_i	Able to lift or carry 10 lbs. at all	29 logical, 53 hot deck	82	3,373	2.43
I37_i	Difficulty using hands or fingers	39 hot deck	39	8,106	0.48
I38_i	Able to use hands or fingers at all	25 logical, 15 hot deck	40	1,966	2.03
I39_i	Difficulty reaching over head	46 hot deck	46	8,106	0.57
I40_i	Able to reach over head at all	35 logical, 21 hot deck	56	1,961	2.86
I41_i	Difficulty standing	75 hot deck	75	8,106	0.93
I42_i	Able to stand at all	31 logical, 21 hot deck	52	4,572	1.11
I43_i	Difficulty stooping	3 logical, 54 hot deck	57	8,106	0.70
I44_i	Able to stoop at all	33 logical, 41 hot deck	74	4,502	1.64
I45_i	Difficulty getting around inside home	34 hot deck	34	8,106	0.42
I46_i	Need help to get around inside home	31 logical, 5 hot deck	36	1,210	2.98
I47_i	Difficulty getting around inside home	7 logical, 33 hot deck	40	8,106	0.49
I48_i	Need help to get around outside home	17 logical, 26 hot deck	43	2,898	1.48
I49_i	Difficulty getting into/out of bed	35 hot deck	35	8,106	0.43
I50_i	Need help getting into/out of bed	21 logical, 17 hot deck	38	2,071	1.84
I51_i	Difficulty bathing or dressing	4 logical, 33 hot deck	37	8,106	0.46
I52_i	Need help bathing or dressing	24 logical, 16 hot deck	40	1,693	2.36
I53_i	Difficulty shopping	15 logical, 31 hot deck	46	8,106	0.57
I54_i	Need help shopping	25 logical, 11 hot deck	36	2,428	1.48

TABLE IV.4 *(continued)*

Variable Name	Description	Imputation Method[a]	Number Missing	Number Eligible	Percent Imputed
I55_i	Difficulty preparing own meals	16 logical, 43 hot deck	59	8,106	0.73
I56_i	Need help to prepare meals	32 logical, 19 hot deck	51	2,607	1.96
I57_i	Difficulty eating	31 hot deck	31	8,106	0.38
I58_i	Need help to eat	28 logical, 8 hot deck	36	937	3.84
I59_i	Trouble concentrating	83 hot deck	83	8,106	1.02
I60_i	Trouble coping with stress	96 hot deck	96	8,106	1.18
I61_i	Trouble getting along with people	78 hot deck	78	8,106	0.96
C_EquipFuncLim_I	Use equipment/device for functional/sensory limitation	16 constructed from imputed variables	16	8,106	0.20
C_NumSenLim_i	Number of sensory limitations	154 constructed from imputed variables	154	8,106	1.90
C_NumSevSenLim_i	Number of severe sensory limitations	46 constructed from imputed variables	46	8,106	0.57
C_NumPhyLim_i	Number of physical functional limitations	242 constructed from imputed variables	242	8,106	2.99
C_NumSevPhyLim_i	Number of severe physical functional limitations	235 constructed from imputed variables	235	8,106	2.99
C_NumEmotLim_i	Number of emotional/social limitations	186 constructed from imputed variables	186	8,106	2.29
C_NumADLs_i	Number of impaired activities of daily living (ADLs)	59 constructed from imputed variables	59	8,106	0.73
C_NumADLAssist_i	Number of ADLs requiring assistance	38 constructed from imputed variables	38	8,106	0.47
C_NumIADLs_i	Number of instrumental activities of daily living (IADL) difficulties	82 constructed from imputed variables	82	8,106	1.01
C_NumIADLAssist_i	Number of IADLs Requiring Assistance	56 constructed from imputed variables	56	8,106	0.69
C_PCS8TOT_i	Physical summary score	279 constructed from imputed variables	279	8,106	3.44

TABLE IV.4 *(continued)*

Variable Name	Description	Imputation Method[a]	Number Missing	Number Eligible	Percent Imputed
C_MCS8TOT_i	Mental summary score	279 constructed from imputed variables	279	8,106	3.44
CageScore_indicator_i	CAGE Alcohol Score	44 constructed from imputed variables	44	8,106	0.54
I72_i	Use drugs in larger amounts than prescribed	73 hot deck	73	8,106	0.90
C_DrugDep_i	Drug dependence	76 constructed from imputed variables	76	8,106	0.94

Source: NBS, round 2.

[a] For all of the imputations using hot deck in this section, a sequential 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 he or she couldn't see newsprint letters at all (I18) or required special devices to read newsprint letters (I19), then I17_i was a 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 other sources 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	98 logical, 22 imputed using specialized procedures	120	8,106	1.48
C_AmtWorkComp_i	Amount received from workers' compensation last month	47 logical, 3 imputed using specialized procedures	50	8,106	0.62
C_AmtVetBen_i	Amount received from veterans' benefits last month	39 logical, 24 imputed using specialized procedures	63	8,106	0.78
C_AmtPubAssis_i	Amount received from public assistance last month	61 logical, 25 imputed using specialized procedures	86	8,106	1.06
C_AmtUnemply_i	Amount received from unemployment benefits last month	47 logical, 5 imputed using specialized procedures	52	8,106	0.64
C_AmtPrivPen_i	Amount received from private pension last month	47 logical, 25 imputed using specialized procedures	72	8,106	0.88
C_AmtOthReg_i	Amount received from other regular sources last month	47 logical, 13 imputed using specialized procedures	60	8,106	0.74

Source: NBS, round 2.

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, the respondent was asked a series of questions on whether the value was above or below specific values. When a respondent could not provide a specific value, he or she 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 into a range of values, 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 he or she received funds from a source, we then imputed whether or not the respondent did 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's 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 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

(FedPovertyLevel_cat1).[34] Also included in this section is the constructed variable for the 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 question in which a specific value was requested, and if a specific value could not be provided, then 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[35] 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.

Logical assignments in this section are based on related variables also in this section. For example, the four logical assignments for L11_i are due to the fact that four respondents did not answer L11, but indicated in L16 that only one adult lived in the household, and in L17 indicated the number of children living with them in the household. For these four respondents, the value for L11 was logically assigned to 1 or 2 depending upon the response to L17.

[34] An additional variable, C_NumChildren_i, was also imputed. This variable is defined as the total number of children in the household plus the number of respondent's children living outside the household. This variable was not used in any subsequent processing and upon further review, was not deemed necessary for analysis, but is in the final file.

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

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	1 logical, 282 hot deck	283	8,106	3.48
L3_i	Highest year/grade completed in school	13 longitudinal imputations; 118 hot deck	131	8,106	1.62
L8_i	Marital Status	52 hot deck	52	8,106	0.64
L11_i	Living arrangements	4 logical, 54 hot deck	58	8,106	0.72
C_NumChildhh_i	Number of children living in the household	3 logical, 43 hot deck	46	8,106	0.57
C_hhsize_i	Household Size	1 logical, 48 hot deck	49	8,106	0.60
C_cohab_i	Cohabitation Status	4 logical, 45 hot deck	49	8,106	0.60
C_FedPovertyLevel_cat1	2004 Federal Poverty Level	2,809 constructed from imputed variables	2,809	8,106	34.65

Source: NBS, round 2.

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 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.

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.[36]

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[37] 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 statisic. 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 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 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 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.

[36] A World Wide Web site 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, as well as links to articles that compare features of the software packages.

[37] 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.[38] 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).

[38] 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

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., F. Potter, and S. 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., G. Livermore, D. Stapleton, J. Kregel, J., T. Silva, B. O'Day, T. Fraker, W. G. Revell Jr., H. Schroeder, and M. Edwards. "Evaluation of the Ticket to Work Program: Initial Evaluation Report" Prepared for the Social Security Administration. Washington DC: Mathematica Policy Research, Inc., 2004.

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

Wright, D., E. Grau, M. Sloan, F. Potter, and K. Barrett."National Beneficiary Survey: Round 2 (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.

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.

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

Question #	Question Text	Current Response Options	Additional Categories Created
B27	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

Question #	Question Text	Current Response Options	Additional Categories Created
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 09= SPECIAL CHAIR / BACK SUPPORT 10= SPECIAL SHOES / SUPPORT STOCKINGS
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 or assistive b=Need changes in {your/NAME's} work c=Need changes to the tasks {you were/NAME was} assigned or how they are performed d=Need changes to the physical work environment e=Need co-workers or others to assist {you/NAME}?

Question #	Question Text	Current Response Options	Additional Categories Created
D23	Why did {you/NAME} stop working at this job?	LAYOFF, FIRED, RETIRED 1=LAYOFF, PLANT CLOSED 2=FIRED 3=RETIRED/OLD AGE 4=JOB WAS TEMPORARY AND ENDED PROBLEMS WITH JOB 5=DID NOT LIKE SUPERVISOR OR CO-WORKERS 6=DID NOT LIKE JOB DUTIES 7=DID NOT LIKE JOB EARNINGS 8=DID NOT LIKE BENEFITS 9=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 government benefits 22=Work schedule
D25b	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 havechosen to work or earn less than {you/he/she} could have during 2004? (SPECIFY: <OPEN>)	g=Had medical problems/complications

Question #	Question Text	Current Response Options	Additional Categories Created
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 such as bathing, dressing, preparing meals, and doing housework? 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 What other special equipment or medical devices?) 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 j=More supportive/helpful employer and/or coworker
E32	Who talked to {you/NAME or his/her representative} about the program?	01=SOCIAL SECURITY ADMINISTRATION 02= MAXIMUS 03= STATE VOCATIONAL REHABILITATION AGENCY, OR {VRNAME} 04= CURRENT/FORMER EMPLOYER 05= FRIEND/FAMILY MEMBER 06= INDEPENDENT LIVING CENTER 07= EMPLOYMENT NETWORK 08=OTHER AGENCY/ORGANIZATION 09=HEALTH CARE PROVIDER 10= OTHER (SPECIFY: <OPEN>)	11=CASE WORKER/SOCIAL WORKER
E37a1	Why {are you/is NAME} no longer receiving services from {EN FROM ROUND 1 E39 OR E46 WHEN E41=01 OR E45=01}?	<OPEN>	1=Never received information/case dropped/ didn't help 2=Found a job 3=I cannot work for health reasons
E43	Why {are you/is NAME} no longer receiving services from {EN IN 2004 FROM E39}?	<OPEN>	01=Never received any info/case dropped/ didn't help 02=Found a job 03=I cannot work for health reasons 04=Other reason related to personal circumstance 05=Other reason related to EN

Question #	Question Text	Current Response Options	Additional Categories Created
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/ couldn't find job 02=Did not know could/did not have ticket 03=Was not healthy enough to participate
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 – NO LONGER THINK JOB OPPORTUNITIES EXIST 07=FEARED SERVICES WOULD ENDANGER BENEFITS 08=INFORMATION TOO CONFUSING – DID NOT KNOW WHERE TO START 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=Already had a job/in school 03=Did not know about/understand the program 04=Did not want to participate 05=Other 06=Cannot work, reason unspecified 07=Did not want to lose benefits/make less money 08=Can't work" responses that do no specify a physical/mental condition
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

Question #	Question Text	Current Response Options	Additional Categories Created
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 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 08=Church or religious institution
G36	In 2004, please tell me if {you/NAME} received any of the following services from {PROVIDER FROM G30_1 DE-DUPLICATED LIST IF USED IN 2004}. Did {you/he/she} receive:	a=Physical therapy? b=Occupational therapy? d=Speech therapy? e=Special equipment or devices? f=Personal counseling or therapy? g=Group therapy? d=Medical services? h=A work or job assessment? i=Help to find a job? j=Training to learn a new job or skill? k=Advice about modifying {your/his/her} job or work place? l=On-the-job training, job coaching, or support services? m=Anything else that I didn't mention? (SPECIFY: <OPEN>)	n=Scholarships/grants/loans

Question #	Question Text	Current Response Options	Additional Categories Created
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 DISABLITIES 10= WORKER'S COMPENSATION 11= DISABILITY INSURANCE 12=OTHER (SPECIFY: <OPEN>)	14=SCHOOL/FINANCIAL AID/GRANT 15=STATE AGENCY/COUNTY/GOVERNMENT
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 PROFESSIONAL 14=COURT/POLICE
G56	How did {your/NAME's} {FILL PERSON(S) FROM G55} pressure {you/him/her} to use these services?	01=SAID {NAME} WOULD LOSE DISABILITY AND/OR HEALTH INSURANCE BENEFITS 02=ENCOURAGED/WOULD NOT TAKE "NO" FOR AN ANSWER 03=THREATENED TO WITHHOLD SERVICES 04=THREATENED TO TAKE AWAY OTHER SUPPORT (E.G., KICK OUT OF THE HOUSE) 05=OTHER (SPECIFY: <OPEN>)	06=THREATENED HOSPITALIZATION/JAIL

Question #	Question Text	Current Response Options	Additional Categories Created
G61	Why {were you/was NAME} unable to get these services?	<OPEN>	01=Not eligible/request refused 02=Lack information on how to get services 03=Could not afford/insurance would not cover 04=Did not try 05=Too difficult/too confusing to get services 06=Problems with the service or agency
H3	Why did {you/NAME} decide to participate in the Ticket to Work program?	<OPEN>	01=Wanted to get a job or more money/benefits 02=Wanted to do something and feel more independent 03=Recommended/told to use it/thought using it was required
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 another agency 02=Already receiving services from VR 03=Didn't understand ticket/didn't know what it was for
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=Not helpful/didn't like 03=Did not like the job offered
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=Information on how and where to use the ticket 02=Information about services provided

Question #	Question Text	Current Response Options	Additional Categories Created
H35	Why did {you/NAME or (his/her) representative} choose {{LONGEST} EMPLOYMENT NETWORK IN 2004}?	01=STAFF WERE MOST RESPONSIVE/COURTEOUS/KNOWLEDGE ABLE 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=Did not receive services needed 03=Problems with counselor 04=Transportation/location problems
H48	What was the problem about?	<OPEN>	01=Trouble making/keeping contact 02=Did not receive services wanted/needed
H50	What did {you/NAME} or someone else do to try to solve the problem?	01=REFERRED TO DOCUMENTS/INFORMATION ABOUT PROVIDER/PROGRAM 02=CONTACTED EN BY PHONE 03=CONTACTED EN IN WRITING 04=CONTACTED PROGRAM MANAGEMENT COMPANY (MAXIMUS) BY PHONE 05=CONTACTED PROGRAM MANAGEMENT COMPANY (MAXIMUS) BY WRITING 06=CONTACTED SSA BY PHONE 07=CONTACTED SSA IN WRITING 08=CONTACTED OTHER STATE/LOCAL AGENCY 09=CONTACTED {LOCAL PROTECTION & ADVOCACY AGENCY} FOR HELP (H) 10=CONTACTED CASE WORKER/JOB COACH 11=OTHER (SPECIFY: <OPEN>)	12=QUIT/LOOKED FOR JOB ON OWN

Question #	Question Text	Current Response Options	Additional Categories Created
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 10=DEVICES TO AIDE IN BREATHING INCLUDING OXYGEN, INHALER, ALBUTEROL, AND/OR NEBULIZER
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

Question #	Question Text	Current Response Options	Additional Categories Created
M2a_rlshp	How are you 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
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

Question #	Question Text	Current Response Options	Additional Categories Created
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
	Healthcare Support
311	Nursing, Psychiatric and Home Health Aides

Code	Occupation
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

TABLE D.1

LOCATION LOGISTIC PROPENSITY MODEL: REPRESENTATIVE BENEFICIARY SAMPLE

Factors in the Location Model		
Main Effects	Parameter Estimate[a]	Standard Error
Number of Moves in Past 5 Years (MOVE_1):		
One, two, three, or four moves	-1.320†	0.449
No moves, old information, or no information about moves	Ref. cell	
Primary Diagnosis Classification (DIG_1):		
Beneficiary had mental disability	0.138†	0.350
Beneficiary had physical disability (excluding deaf cases)	-0.180†	0.386
Beneficiary was deaf, or disability unknown	Ref. cell	
Identity of Payee Relative to Beneficiary (REPREPAYEE_1):		
Beneficiary received benefit payments himself/herself, or from family member	0.693†	0.268
Institution received benefits on behalf of beneficiary, or information unknown	Ref. cell	
Indicator Whether Beneficiary and Applicant for Benefits are in Same Zip Code (PDZIPSAME):		
Applicant and beneficiary live in same zip code	1.312†	0.634
Applicant and beneficiary live in different zip code	0.326†	0.813
Information about whether applicant and beneficiary live in same zip code not given	Ref. cell	
Gender (SEX):		
Female	0.246	0.131
Male	Ref. cell	
Urbanicity of Place of Residence of Beneficiary (METRO_1):		
Beneficiary resides in metropolitan statistical area (MSA)	Ref. cell	
Beneficiary resides in nonmetropolitan area adjacent to large metropolitan area	0.649*	0.311
Beneficiary resides in nonmetropolitan area not adjacent to large metropolitan area	0.356	0.240
Whether Beneficiary is Institutionalized (INSTIT):		
Beneficiary is institutionalized	0.710	0.409
Beneficiary is not institutionalized, or no information	Ref. cell	
Geographic Region (Based on U.S. Census Divisions) of Beneficiary's Place of Residence (DIVISION):		
Pacific	-0.188†	0.315
Mountain	-0.245	0.327
East North Central	-0.312†	0.293
West North Central	0.622	0.401
East South Central	-0.048	0.330
West South Central	0.191†	0.434
South Atlantic	0.02†	0.34
Middle Atlantic	-0.306†	0.302
New England	Ref. cell	
Address of Payee Obtained from SSI File (SSIADDP):		
Yes	-0.384	0.240
No or unknown	Ref. cell	

TABLE D.1 (*continued*)

Factors in the Location Model		
Main Effects	Parameter Estimate[a]	Standard Error
Beneficiary's Living Situation (LIVING_1):		
Beneficiary lives with his or her parents	-1.047**	0.344
Beneficiary does not live with his or her parents, or unknown	Ref. cell	
Number of Phone Numbers on SSA File Over Past Five Years (PHONE_1):		
No record of changes	-0.159†	0.156
One or more changes, or information unknown	Ref. cell	
Beneficiary's Age Category (AGECAT):		
Age in range 18 to 29 years	0.096	0.131
Age in range 30 to 39 years	0.071	0.124
Age in range 40 to 49 years	Ref. cell	
Age in range 50 to 64 years	0.053†	0.199
Two-factor Interactions[b]		
DIVISION * PDZIPSAME		
South Atlantic * Applicant and beneficiary live in same zip code	0.707*	0.299
South Atlantic * Applicant and beneficiary live in different zip codes	-0.806†	0.469
Pacific * Applicant and beneficiary live in different zip codes	-0.697	0.445
Middle Atlantic * Applicant and beneficiary live in different zip codes	1.439*	0.706
West South Central * Applicant and beneficiary live in different zip codes	-1.082†	1.176
DIVISION * DIG_1		
Pacific * Beneficiary had physical disability (excluding deaf cases)	-0.295	0.383
East North Central * Beneficiary had physical disability (excluding deaf cases)	0.753	0.416
South Atlantic * Beneficiary had physical disability (excluding deaf cases)	0.027†	0.375
West South Central * Beneficiary had mental disability	-0.696†	0.498
PDZIPSAME * DIG_1		
Beneficiary had physical disability (excluding deaf cases) * Applicant and beneficiary live in same zip code	0.101	0.562
Beneficiary had physical disability (excluding deaf cases) * Applicant and beneficiary live in different zip codes	0.335†	0.731
Beneficiary had mental disability * Applicant and beneficiary live in same zip code	-0.653	0.541
Beneficiary had mental disability * Applicant and beneficiary live in different zip codes	-0.470†	0.645
PDZIPSAME * REPREPAYEE _1		
Applicant and beneficiary live in same zip code * Beneficiary received benefit payments himself/herself, or from family member	-1.192**	0.395
Applicant and beneficiary live in different zip codes * Beneficiary received benefit payments himself/herself, or from family member	-0.742	0.511
MOVE_1*PHONE_1		
One, two, three, or four moves * No record of phone changes	0.940	0.497
PDZIPSAME * AGECAT		
Applicant and beneficiary live in same zip code * Age category 50 to 64	0.601	0.350
Applicant and beneficiary live in different zip codes * Age category 50 to 64	0.486	0.507

TABLE D.1 (*continued*)

Factors in the Location Model		
Main Effects	Parameter Estimate[a]	Standard Error
Three-factor Interactions[b]		
DIG_1 * DIVISION * PDZIPSAME		
Beneficiary had physical disability (excluding deaf cases) * South Atlantic * Applicant and beneficiary live in different zip codes	1.617	0.913
Beneficiary had mental disability * West South Central * Applicant and beneficiary live in different zip codes	1.871	1.347

[a] Parameter estimates with a cross ([†]) are essentially meaningless because higher order terms that include the variable in question are also in the model. One star (*) and two stars (**) represent significance at the 5 percent and 1 percent levels respectively.

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

TABLE D.2

COOPERATION LOGISTIC PROPENSITY MODEL: REPRESENTATIVE BENEFICIARY SAMPLE

Factors in the Cooperation Model		
Main Effects	Parameter Estimate[a]	Standard Error
Number of Address Changes in the Past Five Years (MOVE_2):		
At most one move	0.928†	0.510
Two or more moves, or information older than five years, or no information	Ref. cell	
Gender (SEX)		
Female	-0.064†	0.248
Male	Ref. cell	
Beneficiary Recipient Benefit Type (SSI_SSDI_2)		
SSDI only	-0.547†	0.265
SSI only, or both SSI and SSDI	Ref. cell	
Disability Diagnosis Classification (DIG):		
Beneficiary had physical disability (excluding deaf cases)	0.682†	0.466
Beneficiary had mental disability	0.352†	0.487
Beneficiary was deaf	0.542†	0.992
Information about disability not given	Ref. cell	
Identity of Payee Relative to Beneficiary (REPREPAYEE_2):		
Family member received benefits on behalf of beneficiary	0.023†	0.269
All other payees (including those with unknown payee identity)	Ref. cell	
Indicator Whether Beneficiary and Applicant for Benefits Are in Same Zip Code (PDZIPSAME_2):		
Applicant and beneficiary live in same zip code	0.197	0.132
Applicant and beneficiary live in different zip code, or no information	Ref. cell	
Urbanicity of Beneficiary's Place of Residence (METRO):		
Beneficiary lived in metropolitan area with population of 1 million or more	Ref. cell	
Beneficiary lived in metropolitan area with population from 250,000 to 1 million	0.148†	0.243
Beneficiary lived in metropolitan area with population under 250,000	-0.225†	0.308
Beneficiary lived in nonmetropolitan area adjacent to large (> 1 million) metropolitan area	-1.630†	0.523
Beneficiary lived in nonmetropolitan area adjacent to metropolitan area under 1 million	0.488†	0.580
Beneficiary lived in nonmetropolitan area not adjacent to metropolitan area	1.549†	0.590
Geographic Region (Based on U.S. Census Divisions) of Beneficiary's Residence (DIVISION_2):		
South Atlantic	0.049	0.130
East North Central	0.115†	0.238
West South Central	0.684†	0.259
All regions except South Atlantic, East North Central, and West South Central	Ref. cell	
Whether the Beneficiary was Hispanic or Not (HISPANICITY):		
Beneficiary was Hispanic	-0.388	0.261
Beneficiary not Hispanic, or unknown	Ref. cell	

TABLE D.2 (*continued*)

Factors in the Cooperation Model		
Main Effects	Parameter Estimate[a]	Standard Error
Race of the Beneficiary (RACE_2):		
White	0.847†	0.524
Asian/Pacific Islander	-0.974†	0.373
Race known to be neither White nor Asian/Pacific Islander, or unknown	Ref. cell	
Beneficiary's Age Category (AGECAT_2):		
Age in range 40 to 49 years	0.131	0.096
Age in range 18 to 39 years, or 50 to 64 years	Ref. cell	
Beneficiary's Type of Claim (TOC_2):		
Disability claim	-0.270†	0.178
Survivor claim, or unknown	Ref. cell	
Number of Phone Numbers on SSA File Over Past Five Years (PHONE_2):		
One or fewer phone changes on SSA file over past five years, or unknown	Ref. cell	
Two or more changes in phone number on SSA file	-1.595†	0.648
Two-factor Interactions[b]		
RACE_2 * METRO		
White * Beneficiary lived in metropolitan area with population from 250,000 to 1 million	-0.644*	0.265
White * Beneficiary lived in metropolitan area with population under 250,000	0.045	0.362
White * Beneficiary lived in nonmetropolitan area adjacent to large (> 1 million) metropolitan area	1.009†	0.565
White * Beneficiary lived in nonmetropolitan area adjacent to metropolitan area under 1 million	-0.986†	0.540
White * Beneficiary lived in nonmetropolitan area not adjacent to metropolitan area	-1.322†	0.943
Asian/Pacific Islander * Beneficiary lived in area that was not a metropolitan area of 1 million or more	-1.786**	0.633
RACE_2 * SSI_SSDI_2		
White * SSDI only	0.187†	0.267
RACE_2 * DIG		
White * Beneficiary had physical disability (excluding deaf cases)	-0.901	0.520
White * Beneficiary had mental disability	-0.561	0.501
White * Beneficiary was deaf	0.443	0.974
RACE_2 * MOVE_2		
White * At most one move	-0.466†	0.349
RACE_2 * GENDER		
White * Female	0.423†	0.265
RACE_2 * DIVISION_2		
White * East North Central	-0.207	0.287

TABLE D.2 (*continued*)

Factors in the Cooperation Model		
Main Effects	Parameter Estimate[a]	Standard Error
DIG * METRO		
Beneficiary had physical disability (excluding deaf cases) * Beneficiary lived in nonmetropolitan area adjacent to metropolitan area of at least 1 million	0.113†	0.457
Beneficiary had physical disability (excluding deaf cases) * Beneficiary lived in a nonmetropolitan area adjacent to metropolitan area under 1 million	-0.106†	0.482
Beneficiary had physical disability (excluding deaf cases) * Beneficiary lived in a nonmetropolitan area not adjacent to metropolitan area	-0.123†	0.554
Beneficiary was deaf * Beneficiary lived in metropolitan area with population from 250,000 to 1 million	-1.324	0.943
Beneficiary was deaf * Beneficiary lived in metropolitan area with fewer than 250,000 people, or in a nonmetropolitan area	0.679	1.283
DIG * MOVE_2		
Beneficiary had physical disability (excluding deaf cases) * At most one move	-0.558†	0.478
Beneficiary had mental disability * At most one move	-1.226†	0.505
Beneficiary was deaf * At most one move	-0.742	0.915
DIG * SSI_SSDI_2		
Beneficiary had mental disability * SSDI only	0.284	0.227
DIG * REPREPAYEE_2		
Beneficiary had mental disability * Family member received benefits on behalf of beneficiary	0.253	0.295
DIG * SEX		
Beneficiary had mental disability * Female	-0.368†	0.242
Beneficiary was deaf * Female	-1.506	0.874
MOVE_2 * SEX		
At most one move * Female	0.520†	0.400
MOVE_2 * METRO		
At most one move * Beneficiary lived in nonmetropolitan area adjacent to large (> 1 million) metropolitan area	1.879†	0.710
At most one move * Beneficiary lived in nonmetropolitan area adjacent to metropolitan area under 1 million	0.718†	0.608
At most one move * Beneficiary lived in nonmetropolitan area not adjacent to metropolitan area	-1.942†	0.983
MOVE_2 * DIVISION_2		
At most one move * West South Central	-0.461	0.385
MOVE_2 * REPREPAYEE_2		
At most one move * Family member received benefits on behalf of beneficiary	0.362	0.254
MOVE_2 * PHONE_2		
At most one move * Two or more changes in phone number on SSA file	1.198	0.723
MOVE_2 * SSI_SSDI_2		
At most one move * SSDI only	-0.573†	0.373

TABLE D.2 (*continued*)

Factors in the Cooperation Model		
Main Effects	Parameter Estimate[a]	Standard Error
METRO * SEX		
Beneficiary lived in nonmetropolitan area adjacent to large (> 1 million) metropolitan area * Female	1.643*	0.746
Beneficiary lived in nonmetropolitan area adjacent to metropolitan area under 1 million * Female	0.716†	0.724
Beneficiary lived in nonmetropolitan area not adjacent to metropolitan area * Female	0.703†	0.751
METRO * PHONE_2		
Beneficiary lived in metropolitan area with population under 250,000, or in nonmetropolitan area adjacent to metropolitan area of 1 million or more * Two or more changes in phone number on SSA file	2.888*	1.324
Beneficiary lived in nonmetropolitan area adjacent to metropolitan area under 1 million, or not adjacent to metropolitan area * Two or more changes in phone number on SSA file	-1.325	0.787
METRO * TOC		
Beneficiary lived in metropolitan area with population from 250,000 to 1 million * Disability claim	0.762**	0.261
Beneficiary lived in metropolitan area with fewer than 250,000 people, or in a nonmetropolitan area * Disability claim	0.230	0.246
Three-factor Interactions[b]		
RACE_2 * MOVE_2 * SEX		
White * At most one move * Female	-0.520	0.443
RACE_2 * METRO * SEX		
White * Beneficiary lived in nonmetropolitan area adjacent to large (> 1 million) metropolitan area * Female	-1.682*	0.849
White * Beneficiary lived in nonmetropolitan area adjacent to metropolitan area under 1 million, or not adjacent to metropolitan area * Female	0.155	0.696
RACE_2 * METRO * MOVE_2		
White * Beneficiary lived in nonmetropolitan area not adjacent to metropolitan area * At most one move	2.161*	0.954
RACE_2 * MOVE_2 * SSI_SSDI_2		
White * At most one move * SSD only	0.915	0.483
DIG * MOVE_2 * SEX		
Beneficiary had mental disability * At most one move * Female	0.818	0.427
DIG * METRO * MOVE_2		
Beneficiary had physical disability (excluding deaf cases) * Beneficiary lived in nonmetropolitan area adjacent to large (> 1 million) metropolitan area * At most one move	-0.807	0.805
Beneficiary had physical disability (excluding deaf cases) * Beneficiary lived in nonmetropolitan area adjacent to metropolitan area under 1 million * At most one move	2.007*	0.821
Beneficiary had physical disability (excluding deaf cases) * Beneficiary lived in nonmetropolitan area not adjacent to metropolitan area * At most one move	1.574	1.047

TABLE D.2 (*continued*)

Factors in the Cooperation Model		
Main Effects	Parameter Estimate[a]	Standard Error
MOVE_2 * METRO * SEX		
At most one move * Beneficiary lived in nonmetropolitan area adjacent to large (> 1 million) metropolitan area* Female	-1.625*	0.788
At most one move * Beneficiary lived in nonmetropolitan area adjacent to metropolitan area under 1 million * Female	-2.025*	0.814
At most one move * Beneficiary lived in nonmetropolitan area not adjacent to metropolitan area * Female	-0.350	0.892

[a] Parameter estimates with a cross (†) are essentially meaningless because higher order terms that include the variable in question are also in the model. One star (*) and two stars (**) represent significance at the 5 percent and 1 percent levels respectively.

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

TABLE D.3

LOCATION LOGISTIC PROPENSITY MODEL: TICKET PARTICIPANT SAMPLE, PHASE 1 MILESTONES AND OUTCOMES[a]

Factors in the Location Model		
Main Effects	Parameter Estimate[b]	Standard Error
Number of Moves in Past 5 Years (MOVE_3):		
No moves	1.587†	0.923
One or more moves	-0.972	0.566
Old information, or no information about moves	Ref. cell	
Whether the Participant was Hispanic or Not (HISPANICITY):		
Participant was Hispanic	-0.065†	0.811
Participant not Hispanic, or unknown	Ref. cell	
Race of the Participant (RACE_3):		
White	1.045†	0.542
Race known not to be white, or unknown	Ref. cell	
Participant's Age Category (AGECAT_3):		
Age in range 18 to 29 years	0.632†	0.757
Age in range 30 to 39 years	-0.742†	0.536
Age in range 40 to 64 years	Ref. cell	
Indicator Whether Participant and Applicant for Benefits Are in Same Zip Code (PDZIPSAME_3):		
Applicant and participant live in different zip code	0.680†	0.974
Applicant and participant live in same zip code, or no information	Ref. cell	
Number of Phone Numbers on SSA File Over Past Five Years (PHONE_3):		
No phone changes on SSA file over past five years	Ref. cell	
One or more changes in phone number on SSA file, or unknown	1.795†	0.630
Participant Recipient Benefit Type (SSI_SSDI_3)		
SSDI only	0.937	0.542
SSI only, or both SSI and SSDI	Ref. cell	
Geographic Region (Based on U.S. Census Regions) of Participant's Residence (REGION_3):		
South	0.007†	0.411
Northeast, Midwest, or West	Ref. cell	
Urbanicity of Participant's Place of Residence (METRO_3):		
Participant lived in metropolitan area with population of 1 million or more	Ref. cell	
Participant did not live in metropolitan area with population of 1 million or more	1.603*	0.699
Whether Participant was Sampled in Round 1 (LONG)		
Participant was not sampled in Round 1	Ref. cell	
Participant was sampled in Round 1	-0.620†	0.437

TABLE D.3 (*continued*)

Factors in the Location Model		
Main Effects	Parameter Estimate[b]	Standard Error
Two-factor Interactions[c]		
MOVE_3 * HISPANICITY		
No moves * Participant was Hispanic	-2.303	1.174
MOVE_3 * PDZIPSAME_3		
No moves * Applicant and participant live in different zip code	-1.662	1.119
PDZIPSAME_3 * REGION_3		
Applicant and participant live in different zip code * South	-2.188*	1.017
AGECAT_3 * LONG		
Age in range 18 to 29 years * Participant sampled in Round 1	-1.324	0.784
Age in range 30 to 39 years * Participant sampled in Round 1	1.587	0.992
RACE_3 * PHONE_3		
White * One or more phone changes on SSA file over past five years, or no information	-2.381**	0.854

[a] If any levels in the main effects are collapsed from all possible levels, the base variable is followed by "_3"
[b] Parameter estimates with a cross (†) are essentially meaningless because higher order terms that include the variable in question are also in the model. One star (*) and two stars (**) represent significance at the 5 percent and 1 percent levels respectively.
[c] All combinations for the listed interactions that are not shown are part of the reference cells

TABLE D.4

COOPERATION LOGISTIC PROPENSITY MODEL: TICKET PARTICIPANT SAMPLE, PHASE 1 MILESTONES AND OUTCOMES[a]

Factors in the Cooperation Model

Main Effects	Parameter Estimate	Standard Error
Number of moves in past 5 years (MOVE_4):		
No moves	3.525†	0.809
One or more moves, old information, or no information about moves	Ref. cell	
Race of the participant (RACE_4):		
White	0.496†	0.620
Black	2.024†	0.746
Race known to be neither white nor black, or unknown	Ref. cell	
Participant's age category (AGECAT_4):		
Age in range 18 to 39 years	-0.991†	0.610
Age in range 40 to 49 years	1.155†	0.705
Age in range 50 to 64 years	Ref. cell	
Indicator whether participant and applicant for benefits are in the same zip code (PDZIPSAME_4):		
Applicant and participant live in the same zip code	-0.289†	0.425
Applicant and participant live in different zip code, or no information	Ref. cell	
Number of phone numbers on SSA file over past five years (PHONE_4):		
No phone changes on SSA file over past five years	Ref. cell	
One or more changes in phone number on SSA file, or unknown	1.438†	0.846
Geographic region (based on U.S. Census regions) of participant's residence (REGION_4):		
South	0.011†	0.299
Northeast, Midwest, or West	Ref. cell	
Disability diagnosis classification (DIG_4):		
Participant had physical disability (excluding deaf cases)	1.676†	0.694
Participant had mental disability	1.209**	0.417
Participant was deaf, or information about disability not given	Ref. cell	
Identity of payee relative to participant (REPREPAYEE_4):		
Participant received benefit payments himself/herself	-0.888†	0.635
Participant did not receive benefit payments himself/herself, or information unknown	Ref. cell	
Participant's living situation (LIVING_4):		
Participant lives alone	-0.657†	0.705
Participant does not live on his or her own, or unknown	Ref. cell	
Whether participant was sampled in Round 1 (LONG)		
Participant was not sampled in Round 1	Ref. cell	
Participant was sampled in Round 1	-0.071†	0.298

TABLE D.4 (*continued*)

Factors in the Cooperation Model		
Main Effects	Parameter Estimate	Standard Error
Two-factor Interactions[a]		
MOVE_4 * LIVING_4		
No moves * Participant lives alone	-2.900**	0.799
PDZIPSAME_4 * RACE_4		
Applicant and participant live in the same zip code * White	0.821†	0.642
PDZIPSAME_4 * DIG_4		
Applicant and participant live in the same zip code * Participant has physical disability (excluding deaf cases)	0.537†	0.649
AGECAT_4 * RACE_4		
Age in range 18 to 39 years * White	-0.263	0.647
Age in range 40 to 49 years * White	-1.817*	0.781
Age in range 18 to 39 years * Black	0.452	0.772
Age in range 40 to 49 years * Black	-2.874**	0.967
AGECAT_4 * REGION_4		
Age in range 18 to 39 years * South	1.591**	0.555
RACE_4 * DIG_4		
White * Participant has physical disability (excluding deaf cases)	1.489†	0.832
Black * Participant has physical disability (excluding deaf cases)	-1.631**	0.609
DIG_4 * MOVE_4		
Participant has physical disability (excluding deaf cases) * No moves	-1.772**	0.634
DIG_4 * LONG		
Participant has physical disability (excluding deaf cases) * Participant sampled in Round 1	0.489†	0.574
RACE_4 * LONG		
White * Participant sampled in Round 1	0.650†	0.511
LIVING_4 * REPREPAYEE_4		
Participant lives alone * Participant received benefit payments himself/herself	2.218**	0.762
PHONE_4 * REPREPAYEE_4		
One or more phone changes on SSA file over past five years, or no information * Participant received benefit payments himself/herself	-2.436*	1.024

TABLE D.4 (*continued*)

Factors in the Cooperation Model		
Main Effects	Parameter Estimate	Standard Error
Three-factor Interactions[c]		
RACE_4 * DIG_4 * PDZIPSAME_4		
White * Participant had physical disability (excluding deaf cases) * Applicant and participant live in the same zip code	-2.540*	1.027
RACE_4 * DIG_4 * LONG		
White * Participant had physical disability (excluding deaf cases) * Participant sampled in Round 1	-2.336*	0.952

[a] If any levels in the main effects are collapsed from all possible levels, the base variable is followed by "_4"
[b] Parameter estimates with a cross ([†]) are essentially meaningless because higher order terms that include the variable in question are also in the model. One star (*) and two stars (**) represent significance at the 5 percent and 1 percent levels respectively.
[c] All combinations for the listed interactions that are not shown are part of the reference cells

TABLE D.5

LOCATION LOGISTIC PROPENSITY MODEL: TICKET PARTICIPANT SAMPLE, PHASE 1 OUTCOMES ONLY[a]

Factors in the Location Model		
Main Effects	Parameter Estimate[b]	Standard Error
Participant's Gender (SEX)		
Male	Ref. cell	
Female	0.874†	0.718
Participant Recipient Benefit Type (SSI_SSDI_5)		
SSDI only	-0.711†	0.663
SSI only, or both SSI and SSDI	Ref. cell	
Number of Phone Numbers on SSA File Over Past Five Years (PHONE_5):		
No phone changes on SSA file over past five years	Ref. cell	
One or more changes in phone number on SSA file, or unknown	1.574**	0.582
Geographic Region (Based on U.S. Census Regions) of Participant's Residence (REGION_5):		
Midwest	-1.563*	0.720
West	-2.752†	0.849
Northeast or South	Ref. cell	
Disability Diagnosis Classification (DIG_5):		
Participant had physical disability (excluding deaf cases)	-1.076	0.584
Participant had mental disability, was deaf, or information about disability not given	Ref. cell	
Urbanicity of Participant's Place of Residence (METRO_5):		
Participant lived in metropolitan area with population of 1 million or more	Ref. cell	
Participant lived in metropolitan area with population from 250,000 to 1 million	2.400*	1.110
Participant did not live in metropolitan area with population of 250,000 or more	0.967	0.536
Participant's Type of Claim (TOC_5):		
Survivor claim	-1.846	1.057
Disability claim, or unknown	Ref. cell	
Two-Factor Interactions[c]		
REGION_5 * SEX		
West * Female	-1.921*	0.928
REGION_5 * SSI_SSDI_5		
West * SSDI only	2.354**	0.727

[a] If any levels in the main effects are collapsed from all possible levels, the base variable is followed by "_5"

[b] Parameter estimates with a cross (†) are essentially meaningless because higher order terms that include the variable in question are also in the model. One star (*) and two stars (**) represent significance at the 5 percent and 1 percent levels respectively.

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

TABLE D.6

COOPERATION LOGISTIC PROPENSITY MODEL: TICKET PARTICIPANT SAMPLE, PHASE 1 OUTCOMES ONLY[a]

Factors in the Cooperation Model		
Main Effects	Parameter Estimate[b]	Standard Error
Participant's Gender (SEX)		
Male	Ref. cell	
Female	0.301	0.218
Address of Payee Obtained From SSI File (SSIADDP):		
Yes	0.394	0.228
No or unknown	Ref. cell	
Participant Recipient Benefit Type (SSI_SSDI_6)		
SSI	-1.205	0.908
Not SSI	Ref. cell	
Participant's Age Category (AGECAT_6):		
Age in range 18 to 39 years	-0.658†	0.218
Age in range 40 to 64 years	Ref. cell	
Urbanicity of Participant's Place of Residence (METRO_6):		
Participant lived in metropolitan area with population of 1 million or more	Ref. cell	
Participant lived in metropolitan area with population from 250,000 to 1 million	-0.641*	0.267
Participant did not live in metropolitan area with population of 250,000 or more	-0.265	0.238
Geographic Region (Based on U.S. Census Regions) of Participant's Residence (REGION):		
South	-0.211†	0.392
Midwest	-0.830†	0.356
West	-0.410†	0.449
Northeast	Ref. cell	
Identity of Payee Relative to Participant (REPREPAYEE_6):		
Participant received benefit payments himself/herself	0.382	0.279
Participant did not receive benefit payments himself/herself, or information unknown	Ref. cell	
Participant's Living Situation (LIVING_6):		
Participant lives alone	1.383	0.936
Participant does not live on his or her own, or unknown	Ref. cell	
Whether Participant was Sampled in Round 1 (LONG)		
Participant was not sampled in Round 1	Ref. cell	
Participant was sampled in Round 1	-0.233†	0.282

TABLE D.6 (*continued*)

Factors in the Cooperation Model		
Main Effects	Parameter Estimate[b]	Standard Error
Two-factor Interactions		
REGION * LONG		
South * Participant was sampled in Round 1	-0.835	0.505
West * Participant was sampled in Round 1	1.301*	0.584
REGION * AGECAT_6		
Midwest * Age in range 18 to 39 years	0.921	0.599

[a] If any levels in the main effects are collapsed from all possible levels, the base variable is followed by "_6"
[b] Parameter estimates with a cross ([†]) are essentially meaningless because higher order terms that include the variable in question are also in the model. One star (*) and two stars (**) represent significance at the 5 percent and 1 percent levels respectively.

TABLE D.7

LOCATION LOGISTIC PROPENSITY MODEL: TICKET PARTICIPANT SAMPLE, PHASE 1 TRADITIONAL[a]

Factors in the Location Model		
Main Effects	Parameter Estimate[b]	Standard Error
Number of Moves in Past 5 Years (MOVE_7):		
One or more moves	-1.516*	0.595
No moves, old information, or no information about moves	Ref. cell	
Participant's Age Category (AGECAT_7):		
Age in range 18 to 49 years	Ref. cell	
Age in range 50 to 64 years	0.535	0.408
Participant's Gender (SEX)		
Male	Ref. cell	
Female	0.539	0.343
Disability Diagnosis Classification (DIG_7):		
Participant had mental disability	0.596	0.288
Participant had physical disability (including deafness), or information about disability not given	Ref. cell	
Participant Recipient Benefit Type (SSI_SSDI_7)		
SSDI	0.554	0.277
Not SSDI	Ref. cell	

[a] If any levels in the main effects are collapsed from all possible levels, the base variable is followed by "_7"
[b] Parameter estimates with one star (*) and two stars (**) represent significance at the 5 percent and 1 percent levels respectively.

TABLE D.8

COOPERATION LOGISTIC PROPENSITY MODEL: TICKET PARTICIPANT SAMPLE, PHASE 1 TRADITIONAL[a]

Factors in the Cooperation Model		
Main Effects	Parameter Estimate[b]	Standard Error
Number of Moves in Past 5 Years (MOVE_8):		
No moves	0.979†	0.395
One or more moves, old information, or no information about moves	Ref. cell	
Race of the Participant (RACE_8):		
White	0.726*	0.329
Black	0.764*	0.324
Race known to be neither white nor black, or unknown	Ref. cell	
Participant's Gender (SEX)		
Male	Ref. cell	
Female	-0.478*	0.194
Whether the Participant was Hispanic or Not (HISPANICITY):		
Participant was Hispanic	0.927	0.491
Participant not Hispanic, or unknown	Ref. cell	
Indicator Whether Participant and Applicant for Benefits Are in Same Zip Code (PDZIPSAME_8):		
Applicant and participant live in same zip code	-0.243†	0.263
Applicant and participant live in different zip code, or no information	Ref. cell	
Participant's Type of Claim (TOC_8):		
Survivor claim	1.315*	0.505
Disability claim, or unknown	Ref. cell	
Geographic Region (Based on U.S. Census Regions) of Participant's Residence (REGION_8):		
West	1.226*	0.499
Northeast	-0.397	0.212
Midwest or South	Ref. cell	
Disability Diagnosis Classification (DIG_8):		
Participant had physical disability (excluding deaf cases)	0.957†	0.441
Participant had mental disability	0.376	0.300
Participant was deaf, or information about disability not given	Ref. cell	
Identity of Payee Relative to Participant (REPREPAYEE_8):		
Family member received benefits on behalf of participant	-0.552	0.318
All other payees (including those with unknown payee identity)	Ref. cell	
Urbanicity of Participant's Place of Residence (METRO_8):		
Participant lived in metropolitan area with population of 1 million or more	Ref. cell	
Participant lived in metropolitan area with population from 250,000 to 1 million	0.231†	0.271
Participant lived in metropolitan area with population under 250,000	-0.287†	0.559
Participant lived in nonmetropolitan area	-0.909†	1.195

TABLE D.8 (*continued*)

Factors in the Cooperation Model		
Main Effects	Parameter Estimate[b]	Standard Error
Address of Payee Obtained From SSI File (SSIADDP):		
Yes	0.710	0.575
No or unknown	Ref. cell	
Participant Recipient Benefit Type (SSI_SSDI)		
SSDI Only	-0.267	0.334
SSI Only	Ref. cell	
Both SSI and SSDI	0.719†	0.333
Two-factor Interactions[c]		
MOVE_8 * DIG_8		
No moves * Participant has physical disability (excluding deaf cases)	-1.733**	0.608
MOVE_8 * METRO_8		
No moves * Participant lived in metropolitan area with population from 250,000 to 1 million)	1.202**	0.420
No moves * Participant lived in metropolitan area with population under 250,000	-2.041	1.095
No moves * Participant lived in nonmetropolitan area	0.485	1.461
SSIADDP * METRO_8		
Address of payee obtained from SSI file * Participant did not live in metropolitan area over 250,000	1.435	1.208
PDZIPSAME_8 * DIG_8		
Applicant and participant live in same zip code * Participant has physical disability (excluding deaf cases)	1.680*	0.716
SSI_SSDI * DIG_8		
Participant received both SSI and SSDI * Participant has physical disability (excluding deaf cases)	-1.644**	0.398

[a] If any levels in the main effects are collapsed from all possible levels, the base variable is followed by "_7"
[b] Parameter estimates with one star (*) and two stars (**) represent significance at the 5 percent and 1 percent levels respectively.
[c] All combinations for the listed interactions that are not shown are part of the reference cells

TABLE D.9

LOCATION LOGISTIC PROPENSITY MODEL: TICKET PARTICIPANT SAMPLE, PHASE 2 MILESTONES AND OUTCOMES[a]

Factors in the Location Model		
Main Effects	Parameter Estimate[b]	Standard Error
Participant's age category (AGECAT_9):		
Age in range 18 to 29 years	-1.751†	0.834
Age in range 30 to 39 years	-1.226	0.647
Age in range 40 to 64 years	Ref. cell	
Participant's gender (SEX)		
Female	0.627†	0.554
Male	Ref. cell	
Indicator whether participant and applicant for benefits are in the same zip code (PDZIPSAME_9):		
Applicant and participant live in the same zip code	1.903†	0.537
Applicant and participant live in different zip code, or no information	Ref. cell	
Geographic region (based on U.S. Census regions) of participant's residence (REGION_9):		
South, West	5.232†	1.181
Northeast, Midwest	Ref. cell	
Disability diagnosis classification (DIG_9):		
Participant had mental disability	-1.481**	0.473
Participant had physical disability (including deafness), or information about disability not given	Ref. cell	
Identity of payee relative to participant (REPREPAYEE_9):		
Family member received benefits on behalf of participant	-0.170†	0.751
All other payees (including those with unknown payee identity)	Ref. cell	
Participant's living situation (LIVING_9):		
Participant lives alone	-1.303*	0.540
Participant does not live on his or her own, or unknown	Ref. cell	
Urbanicity of place of residence of beneficiary (METRO_9):		
Beneficiary resides in metropolitan statistical area (MSA)	Ref. cell	
Beneficiary resides in nonmetropolitan area adjacent to large metropolitan area	-1.419**	0.509
Beneficiary resides in nonmetropolitan area not adjacent to large metropolitan area	-1.052	0.583

TABLE D.9 (*continued*)

Factors in the Location Model		
Main Effects	Parameter Estimate[b]	Standard Error
Two-factor Interactions[c]		
REPREPAYEE_9 * AGECAT_9 Family member received benefits on behalf of participant * Age in range 18 to 29 years	2.932*	1.393
REGION_9 * SEX West, South * Female	-2.903**	0.882
REGION_9 * PDZIPSAME_9 West, South * Applicant and participant live in same zip codes	-3.480**	0.855

[a] If any levels in the main effects are collapsed from all possible levels, the base variable is followed by "_9"
[b] Parameter estimates with a cross (†) are essentially meaningless because higher order terms that include the variable in question are also in the model. One star (*) and two stars (**) represent significance at the 5 percent and 1 percent levels respectively.
[c] All combinations for the listed interactions that are not shown are part of the reference cells

TABLE D.10

COOPERATION LOGISTIC PROPENSITY MODEL: TICKET PARTICIPANT SAMPLE, PHASE 2 MILESTONES AND OUTCOMES[a]

Factors in the Cooperation Model		
Main Effects	Parameter Estimate[b]	Standard Error
Participant's gender (SEX)		
Male	Ref. cell	
Female	-0.574	0.347
Participant's type of claim (TOC_10):		
Disability claim	-1.235*	0.495
Survivor claim, or unknown	Ref. cell	
Geographic region (based on U.S. Census regions) of participant's residence (REGION_10):		
South, West	-1.479†	0.631
Northeast, Midwest	Ref. cell	
Disability diagnosis classification (DIG_10):		
Participant had physical disability (excluding deaf cases)	2.266†	1.182
Participant had mental disability, was deaf, or information about disability not given	Ref. cell	
Identity of payee relative to participant (REPREPAYEE):		
Participant received benefit payments himself/herself	1.623†	0.722
Family member received benefits on behalf of participant	0.910	0.690
All other payees (including those with unknown payee identity)	Ref. cell	
Participant's living situation (LIVING_10):		
Participant lives alone	-2.195†	0.725
Participant does not live on his or her own, or unknown	Ref. cell	
Number of phone numbers on SSA file over past five years (PHONE_10):		
No phone changes on SSA file over past five years	Ref. cell	
One or more changes in phone number on SSA file, or unknown	-0.860*	0.430
Two-factor Interactions[c]		
REPREPAYEE * DIG_10		
Participant received benefit payments himself/herself * Participant had physical disability (excluding deaf cases)	-2.097	1.144
LIVING_10 * REGION_10		
Participant lives alone * West, South	2.043*	0.840

[a] If any levels in the main effects are collapsed from all possible levels, the base variable is followed by "_10"

[b] Parameter estimates with a cross (†) are essentially meaningless because higher order terms that include the variable in question are also in the model. One star (*) and two stars (**) represent significance at the 5 percent and 1 percent levels respectively.

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

TABLE D.11

LOCATION LOGISTIC PROPENSITY MODEL: TICKET PARTICIPANT SAMPLE, PHASE 2 OUTCOMES ONLY[a]

Factors in the Location Model		
Main Effects	Parameter Estimate[b]	Standard Error
Participant's gender (SEX)		
Male	Ref. cell	
Female	2.476†	1.223
Race of the participant (RACE_11):		
White	-1.349	0.916
Race known not to be white, or unknown	Ref. cell	
Participant's type of claim (TOC_11):		
Disability claim	-2.110	1.241
Survivor claim, or unknown	Ref. cell	
Disability diagnosis classification (DIG_11):		
Participant had physical disability (excluding deaf cases)	1.644†	1.117
Participant was deaf	-2.490**	0.716
Participant had mental disability, or information about disability not given	Ref. cell	
Geographic region (based on U.S. Census regions) of participant's residence (REGION_11):		
South	1.577*	0.693
Northeast, Midwest, or West	Ref. cell	
Participant recipient benefit type (SSI_SSDI_11)		
SSDI Only	1.380*	0.635
SSI Only, or Both SSI and SSDI	Ref. cell	
Urbanicity of participant's place of residence (METRO_11):		
Participant lived in metropolitan area with population of 1 million or more	Ref. cell	
Participant lived in metropolitan area with population from 250,000 to 1 million	1.516*	0.633
Participant lived in did not live in metropolitan area with at least 250,000 population	1.651	1.064
Two-factor Interactions[c]		
SEX * DIG_11		
Female* Participant had physical disability (excluding deaf cases)	-4.076*	1.765

[a] If any levels in the main effects are collapsed from all possible levels, the base variable is followed by "_11"

[b] Parameter estimates with a cross (†) are essentially meaningless because higher order terms that include the variable in question are also in the model. One star (*) and two stars (**) represent significance at the 5 percent and 1 percent levels respectively.

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

TABLE D.12

COOPERATION LOGISTIC PROPENSITY MODEL: TICKET PARTICIPANT SAMPLE, PHASE 2 OUTCOMES ONLY[a]

Factors in the Cooperation Model		
Main Effects	Parameter Estimate[b]	Standard Error
Participant's gender (SEX)		
Male	Ref. cell	
Female	1.213†	0.673
Race of the participant (RACE_12):		
White	-0.989†	0.451
Race known not to be white, or unknown	Ref. cell	
Indicator whether participant and applicant for benefits are in same zip code (PDZIPSAME_12):		
Applicant and participant live in same zip code	0.893†	0.317
Applicant and participant live in different zip code, or no information	Ref. cell	
Disability diagnosis classification (DIG_12):		
Participant had physical disability (excluding deaf cases)	1.538**	0.427
Participant had mental disability	0.585†	0.542
Participant was deaf, or information about disability not given	Ref. cell	
Geographic region (based on U.S. Census regions) of participant's residence (REGION_12):		
South	-0.104†	0.522
Northeast	-0.688	0.416
Midwest, West	Ref. cell	
Participant recipient benefit type (SSI_SSDI_12)		
SSDI	0.341†	0.512
SSI Only	Ref. cell	
Two-factor Interactions[c]		
SEX * SSI_SSDI_12		
Female* SSDI	-1.221	0.740
RACE_12 * DIG_12		
White * Participant had mental disability	0.925	0.552
REGION_12 * PDZIPSAME_12		
South * Applicant and participant live in same zip code	-1.269*	0.625

[a] If any levels in the main effects are collapsed from all possible levels, the base variable is followed by "_12"

[b] Parameter estimates with a cross (†) are essentially meaningless because higher order terms that include the variable in question are also in the model. One star (*) and two stars (**) represent significance at the 5 percent and 1 percent levels respectively.

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

TABLE D.13

LOCATION LOGISTIC PROPENSITY MODEL: TICKET PARTICIPANT SAMPLE, PHASE 2 TRADITIONAL[a]

Factors in the Location Model		
Main Effects	Parameter Estimate[b]	Standard Error
Participant's age category (AGECAT_13):		
Age in range 18 to 29 years	-0.113	0.404
Age in range 30 to 39 years	-1.153*	0.412
Age in range 40 to 64 years	Ref. cell	
Race of the participant (RACE_13):		
Black	-0.936**	0.316
Race known not to be black, or unknown	Ref. cell	
Disability diagnosis classification (DIG_13):		
Participant was deaf	-1.991**	0.654
Participant had mental disability	-0.619	0.369
Participant had physical disability (excluding deafness), or information about disability not given	Ref. cell	
Urbanicity of participant's place of residence (METRO_13):		
Participant lived in metropolitan area with population of 250,000 or more	Ref. cell	
Participant lived in metropolitan area with population under 250,000	-1.051*	0.427
Participant lived in nonmetropolitan area	0.204	0.485
Participant recipient benefit type (SSI_SSDI_13)		
Both SSI and SSDI	-0.670	0.557
SSI Only or SSDI Only	Ref. cell	
Participant's living situation (LIVING_13):		
Participant lives alone	1.423	0.728
Participant does not live on his or her own, or unknown	Ref. cell	

[a] If any levels in the main effects are collapsed from all possible levels, the base variable is followed by "_13"
[b] Parameter estimates with one star (*) and two stars (**) represent significance at the 5 percent and 1 percent levels respectively.

TABLE D.14

COOPERATION LOGISTIC PROPENSITY MODEL: TICKET PARTICIPANT SAMPLE, PHASE 2 TRADITIONAL[a]

Factors in the Cooperation Model		
Main Effects	Parameter Estimate[b]	Standard Error
Participant's age category (AGECAT_14):		
Age in range 18 to 29 years	-0.369†	0.654
Age in range 30 to 39 years	-1.644†	0.283
Age in range 40 to 64 years	Ref. cell	
Indicator whether participant and applicant for benefits are in same zip code (PDZIPSAME):		
Applicant and participant live in same zip code	0.760†	0.481
Applicant and participant live in different zip code	-0.906	0.516
No information about whether applicant and participant live in same zip code	Ref. cell	
Disability diagnosis classification (DIG_14):		
Participant had physical disability, including deafness	0.465†	0.321
Participant had mental disability, or information about disability not given	Ref. cell	
Geographic region (based on U.S. Census regions) of participant's residence (REGION_14):		
Northeast, South	1.286†	0.606
Midwest	1.115**	0.312
West	Ref. cell	
Number of moves in past 5 years (MOVE_14):		
No moves	0.678	0.336
One or more moves, old information, or no information about moves	Ref. cell	
Participant's type of claim (TOC_14):		
Disability claim	-0.552†	0.517
Survivor claim, or unknown	Ref. cell	
Urbanicity of participant's place of residence (METRO_14):		
Participant lived in metropolitan area with population of 250,000 or more	Ref. cell	
Participant did not live in metropolitan area with population of 250,000 or more	0.067†	0.491
Two-factor Interactions[c]		
AGECAT_14 * REGION_14		
Age within range 18 to 29 years * Northeast, South	-0.229	0.803
Age within range 30 to 39 years * Northeast, South	2.225**	0.699
METRO_14 * TOC_14		
Participant did not live in metropolitan area with population of 250,000 or more * Disability claim	1.585*	0.737

TABLE D.14 (*continued*)

Factors in the Cooperation Model		
Main Effects	Parameter Estimate[b]	Standard Error
DIG_14 * PDZIPSAME		
Participant had physical disability, including deafness * Applicant and participant live in same zip code	-1.813**	0.511

[a] If any levels in the main effects are collapsed from all possible levels, the base variable is followed by "_14"

[b] Parameter estimates with a cross ([†]) are essentially meaningless because higher order terms that include the variable in question are also in the model. One star (*) and two stars (**) represent significance at the 5 percent and 1 percent levels respectively.

[c] 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 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:	**Wtr2_ben**
Participant sample:	**Wtr2_par**
Combined samples:	**Wgt1_Comb**

Nest Variables

A_STRATA

1. Clustered samples for both beneficiaries and participants
   a. A_STRATA = 1000 for PSUs in Phase 1 states
   b. A_STRATA = 2000 for PSUs in Phase 2 states
   c. A_STRATA = 3000 for PSUs in Phase 3 states

2. Unclustered samples for participants requiring unclustered sample
   a. A_STRATA = 1111 Outcome-only participants in PSUs in Phase 1 states, R 1 frame
   b. A_STRATA = 1121 Outcome-only participants not in PSUs in Phase 1 states, R1 frame
   c. A_STRATA = 1112 Outcome-only participants in PSUs in Phase 1 states, R 2 frame
   d. A_STRATA = 1122 Outcome-only participants not in PSUs in Phase 1 states, R2 frame
   e. A_STRATA = 1211 Milestones and outcomes participants in PSUs in Phase 1 states, R 1 frame

f. A_STRATA = 1221 Milestones and outcomes participants not in PSUs in Phase 1 states, R1 frame
g. A_STRATA = 1212 Milestones and outcomes participants in PSUs in Phase 1 states, R 2 frame
h. A_STRATA = 1222 Milestones and outcomes participants not in PSUs in Phase 1 states, R2 frame
i. A_STRATA = 2112 Outcome-only participants in PSUs in Phase 2 states, R 1 frame
j. A_STRATA = 2122 Outcome-only participants not in PSUs in Phase 2 states, R1 frame
k. A_STRATA = 2212 Milestones and outcomes participants in PSUs in Phase 2 states, R 2 frame
l. A_STRATA = 2222 Milestones and outcomes participants not in PSUs in Phase 1 states, R2 frame

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 Milestones and outcomes or Outcome-only participants

Notes:

1. Before each SUDAAN procedure, sort by A_STRATA and A_PSU

2. Use SUDAAN's SUBPOPN statement to define population for which estimates are wanted.

   For example, for estimates of SSI participant population, use SUBPOPN to define SSI participants.