National Beneficiary Survey Round 4

(Volume 1 of 3): Editing, Coding, Imputation, and Weighting Procedures

February 3, 2012

Eric Grau
Kirsten Barrett
Debra Wright
Yuhong Zheng
Barbara Carlson
Frank Potter
Sara Skidmore

MATHEMATICA
Policy Research

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

Contract Number:
0600-03-60129

OMB Number:
0960-0666
OMB Expiration Date:
3/31/2013

Mathematica Reference Number:
08978.421

Submitted to:
Social Security Administration
Office of Retirement and
Disability Policy
ITC Building
500 E. St., SW, 9th Floor
Washington, DC 20254
Project Officer: Paul O'Leary

Submitted by:
Mathematica Policy Research
1100 First Street, NE, 12th Floor
Washington, DC 20002-4221
Telephone: (202) 484-9220
Facsimile: (202) 863-1763
Project Director: Debra Wright

National Beneficiary Survey Round 4

(Volume 1 of 3): Editing, Coding, Imputation, and Weighting Procedures

February 3, 2012

Eric Grau
Kirsten Barrett
Debra Wright
Yuhong Zheng
Barbara Carlson
Frank Potter
Sara Skidmore

MATHEMATICA
Policy Research

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

ERRATA

(Updated December 20, 2016)

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

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

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

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

CONTENTS

TABLES

ACRONYMS

AIC	Akaike's Information Criterion
CAPI	Computer Assisted Personal Interviewing
CATI	Computer Assisted Telephone Interviewing
CHAID	Chi-Squared Automatic Interaction Detector
CR	Cost Reimbursement Provider Payment Program
EN	Employment Networks
ICD-9	International Classification of Diseases, 9th Revision
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.)
SVRA	State Vocational Rehabilitation Agency
TRS	Telecommunications Relay Service
TTW	Ticket to Work
TTY	Teletypewriter

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

I. INTRODUCTION

As part of an evaluation of the Ticket to Work and Self-Sufficiency program (TTW), Mathematica Policy Research conducted Round 4 of the National Beneficiary Survey (NBS) in 2010. The survey, sponsored by the Social Security Administration's (SSA) Office of Retirement and Disability Policy, 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). Mathematica collected the data by using computer-assisted telephone interviewing (CATI), along with computer-assisted personal interviewing (CAPI) followup for CATI nonrespondents and those preferring or needing an in-person interview to accommodate their disabilities.

A voluntary employment program for people with disabilities, TTW was authorized by the Ticket to Work and Work Incentives Improvement Act of 1999 (TTWIIA). The legislation was designed to create market-driven services to help disability beneficiaries become economically self-sufficient. Under the program, 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 the beneficiary's choice (called Employment Networks, or ENs).[1]

A. 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 participation analysis. Along with the NBS, 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 Supplemental Security Income (SSI) and Social Security Disability Insurance (SSDI) beneficiaries, particularly the 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 on beneficiaries who do not use their Ticket and the reasons behind their decision.

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 beneficiaries' perceptions about TTW and other SSA programs designed to help SSA beneficiaries with disabilities find and keep jobs.

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

In addition to the meeting the original study objectives, the Round 4 survey assessed the impact of changes made to the TTW program in July 2008, when new regulations took effect.

Round 4 NBS data will be combined with SSA administrative data to provide critical information on access to jobs and employment outcomes for beneficiaries, including those who do and do not participate in the TTW program. Though some sections of the NBS target beneficiary activity directly related to TTW, most of the survey captures general information on SSA beneficiaries, including their disability, interest in work, use of services, and employment. As a result, SSA and external researchers interested in disability and employment issues may use the survey data for other policymaking and program-planning efforts.

B. Sample Design Overview

SSA implemented the TTW program in three phases over three years, with each phase corresponding to about one-third of the states. The initial NBS design called for four 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 introduced (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. The survey of all beneficiaries is referred to as the Representative Beneficiary Sample, and the cross-sectional and longitudinal surveys of Ticket participants are referred to as the Ticket Participant Sample. The design subsequently underwent revision to accommodate Phase 1 data collection starting in 2004 rather than in 2003. In addition, Round 4 was postponed until 2010 to address the experiences of TTW participants under the new TTW regulations that took effect in July 2008. In Table I.1, we provide the original planned sample sizes for all rounds of data collection.

Under the initial design, the Round 4 surveys were to concentrate largely on following the Ticket Participant Sample interviewed in earlier rounds and on interviewing new Ticket participants in Phase 3 states. The cross-sectional Representative Beneficiary Sample in Round 4 was to be substantially smaller than the cross-sections in earlier rounds. However, changes in the Federal regulations that substantially altered the TTW program made it less meaningful to track the long-term experiences of beneficiaries who participated in the program under the old regulations. As a result, Ticket participants from previous rounds were not re-interviewed in Round 4 as part of the longitudinal sample and the sample design underwent revision to include a larger cross-section sample of beneficiaries and a representative cross-sectional Ticket Participant Sample.

---

[2] The Ticket to Work program, implemented in 2002, was phased in nationwide over three years. In 2002, the first year of the program, SSA distributed Tickets in 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. Phase 2 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, and Virginia. Phase 3 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—Initial Design

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

[a] Sample sizes refer to number of completed interviews.

[b] (1) = TTW participant longitudinal sample; (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 up to three times in these counts, depending on the number of completed interviews for the sample member in question.

In Rounds 1 through 3, we stratified Ticket participants by the implementation phase of their state of residence and, within each phase, according to the reimbursement system under which their Ticket provider received payments: the traditional cost reimbursement (CR) program, the milestone-outcome payment system, or the outcome-only payment system.[3] In the fourth round, it was no longer necessary to stratify by implementation phase since the TTW program was up and running in all areas. In Rounds 1 through 3, many of the Ticket participants sampled as having a Ticket assigned to a milestone-outcome or outcomes-only provider were signed up with State Vocational Rehabilitation Agencies (SVRA) rather than with ENs. Thus, the first three rounds overrepresented participants signed up with SVRAs. To compensate, in Round 4, we stratified the participant sample by the following provider and payment types: (1) participants with Tickets assigned to SVRAs receiving payments under the traditional CR payment system (referred to in this report as "traditional SVRA") and (2) participants with Tickets assigned to ENs or SVRAs functioning as ENs under the TTW program (referred to in this report as "non–SVRA ENs" and "SVRA ENs"). Participants who assigned their Ticket to an EN were oversampled. Because the number of tickets assigned to the SVRA ENs and Non-SVRA ENs was low among Ticket participants, we selected both a clustered and unclustered sample of participants for each provider type. The sample of participants using the traditional payment type was limited to a clustered sample. The target number

[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: (1) an outcome-only payment system or (2) a milestone-outcome payment system. Under both systems, SSA makes up to 60 monthly payments to the EN for each assigned beneficiary who does not receive 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 it will also pay the EN for up to four milestones achieved by a beneficiary.

of completed interviews for participants in the cross-sectional samples in Round 4 was 3,000 overall, with a target of approximately 750 interviews each for traditional SVRAs and SVRA ENs and 1,500 interviews for non–SVRA ENs.

As in prior rounds, we stratified the cross-sectional Representative Beneficiary Sample by four age-based strata: 18- to 29-year-olds, 30- to 39-year-olds, 40- to 49-year-olds, and 50-year-olds and older. To ensure a sufficient number of persons seeking work, beneficiaries in the first three cohorts were oversampled (18- to 49-year-olds). The target number of completed interviews for Round 4 was 667 beneficiaries in each of the three younger age groups (18 to 29 years, 30 to 39 years, and 40 to 49 years). For those 50 years and older, the target number of completed interviews totaled 400 beneficiaries. Table 1.2 summarizes the actual sample sizes and number of completed interviews for both samples under the revised design.

Table I.2. Round 4 Actual Sample Sizes, Target Completes, and Completes—Implemented Design

Sampling Strata	Sample Size	Target Completed Interviews	Actual Completed Interviews
Representative Beneficiary Sample	3,683	2,400	2,298
Age 18 to 29	1,029	666	634
Age 30 to 39	1,032	666	625
Age 40 to 49	1,019	666	643
Age 50 and older	603	402	396
Ticket Participant Sample	4,334	3,000	2,780
Employment Networks	3,251	2,250	2,030
Non–SVRA providers	2,157	1,500	1,352
SVRA providers	1,094	750	678
Traditional SVRA	1,083	750	750
Total Sample Size	8,017	5,400	5,078

Source: NBS, Round 4.

For all survey rounds, the NBS used a multistage sampling design with a supplemental single-stage sample for some Ticket participant populations. For the multistage design, data from SSA on the counts of eligible beneficiaries in each county formed the primary sampling units (PSU), which consisted of one or more counties. The same PSUs were used for all four rounds, with the selection of PSUs occurring in Round 1. The sampling design section of the User's Guide (Wright et al. 2012) details the selection of PSUs.

C. Round 4 Survey Overview

The NBS was designed and implemented to maximize both response and data quality. Table 1.3 describes the most significant sources of potential non-sampling error identified at the outset of the NBS and describes the ways we attempted to minimize the impact of each. A more detailed discussion of our approach to minimizing total survey error can be found in Appendix A of the Round 4 User's Guide (Wright et al. 2012).

Table I.3. Sources of Error, Description, and Methods to Minimize Impact

Sources of Error	Description	Effort of Minimized
Specification	Error that results when the concept intended to be measured by the question is not the same as the concept the respondent ascribes to the question.	Cognitive interviewing during survey development[4] and pretesting; use of proxy if sample member unable to respond due to cognitive disability
Unit Non-response	Error that occurs when selected sample member is unwilling or unable to participate (failure to interview). Can result in increased variance and potential for bias in estimates if non-responders have different characteristics than responders.	Interviewer training; intensive locating; in-person data collection; refusal conversion; incentives; non-response adjustment to weights.
Item Non-response	Error that results when items are left blank or the respondent reports that he/she does not know the answer or refuses to provide an answer (failure to obtain and record data for all items). Can result in increased variance and potential bias in estimates if non-responders have different characteristics than responders.	Use of probes; allowing for variations in reporting units; assurance of confidentiality; assistance during interview; use of proxy if sample member unable to respond due to cognitive disability; imputation on key variables.
Measurement Error	Errors that occur as a result of the respondent or interviewer providing incorrect information (either intentionally or unintentionally). May result from inherent differences in interview mode.	Same instrument used in both interview modes; Use of probes; adaptive equipment; interviewer training, validation of in-field interviews; assistance during interview; use of proxy if sample member unable to respond due to cognitive disability
Data Processing Errors	Errors in data entry, coding, weighting, and/or analyses.	Coder training; monitoring and quality control checks of coders; quality assurance review of all weighting and imputation procedures

Item non-response was not expected to be a large source of error since there were few obviously sensitive items in the survey. In fact, item non-response was greater than 5 percent only for select items asking for wages and household income. Unit non-response was the greater concern given the population, thus the survey was designed to be executed as a dual-mode survey. Mathematica made initial attempts to interview beneficiaries using CATI followed by CAPI of nonrespondents. CAPI interviews were attempted with respondents who requested an in- person interview, needed an in-person interview to accommodate a disability, or did not have a telephone or

[4] Conducted during survey development phase under a separate contract held by Westat.

whose telephone number could not be located.[5] If a sample person was not able to participate in the survey because of his or her disability, Mathematica sought a proxy respondent. 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. To reduce measurement error, the survey instrument was identical in each mode.

Round 4 CATI data collection for both Representative Beneficiary and Ticket Participant samples began in April 2010.[6] Beginning in August 2010, Mathematica began in-person locating and CAPI and continued, concurrent with CATI interviewing, through December 2010. As shown in Table I.2, the NBS Round 4 sample comprised 3,683 cases selected for the Representative Beneficiary Sample and 4,334 cases selected for the Ticket Participant Sample (for a total of 8,017 cases).[7]

1. Completes and Response Rates

In total, Mathematica completed 5,078 cases (including 38 partially completed interviews)[8]—2,298 from the Representative Beneficiary Sample and 2,780 from the Ticket Participant Samples. An additional 222 beneficiaries and 77 Ticket participants were deemed ineligible for the survey.[9] Across both samples, Mathematica completed 3,936 cases by telephone and 1,142 by CAPI. Proxy interviews were completed for 998 sample members. In 152 cases, the sample member was unable to participate and a proxy could not be identified. The weighted response rate for the Representative Beneficiary Sample was 72.8 percent. The weighted response rate for the Ticket Participant Sample was 71.4 percent. More information about sample selection and sampling weights is available in Grau et al. (2012).

2. Nonresponse Bias

Because the weighted response rates within strata ranged from 67.3 to 75.2 percent and the overall response rate was less than 80 percent, we conducted a nonresponse bias analysis at the conclusion of data collection using all 8,017 sample cases, to determine if there were systematic differences between respondents and nonrespondents that could result in nonresponse bias. This analysis was not conducted in previous rounds of the NBS, since the response rates were close to or exceeded 80 percent (the assumption was that that the effect of nonresponse bias on final estimates was minimal).

[5] Ticket participants in the unclustered sample were not eligible for in-person interviewing.

[6] Interviewing began approximately eight months after the sample was selected.

[7] Given that the clustered and unclustered samples of the Ticket Participant Sample were independent, it was possible 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 duplicate cases were conducted only once but recorded twice (once for each sample). The counts given above include the duplicates as separate cases.

[8] Partial interviews were considered as completed if responses were provided through Section H of the interview (or, if the respondent was not eligible to receive Section H, through Section G).

[9] Ineligible sample members include those who were deceased, incarcerated, and 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 the completion of sampling.

In sum, our analysis indicates that the nonresponse adjustment alleviated nearly all differences observed between respondents and nonrespondents in both the beneficiary and participant samples with two exceptions for the beneficiary sample. First, the nonresponse-adjusted weighted estimate of the beneficiary type differed from the frame for SSI-only cases, even though the original estimate (including all sampled cases) did not differ from the frame. Second, the nonresponse-adjusted weighted proportion of Asians is significantly less than the frame value. The full nonresponse bias analysis can be obtained from SSA (http://www.socialsecurity.gov/disabilityresearch/publicusefiles.html).

D. Data Documentation Reports

The following publically available reports are available from SSA on their website (http://www.socialsecurity.gov/disabilityresearch/publicusefiles.html):

- **User's Guide for Restricted- and Public-Use Data Files** (Wright et al. 2012). This report provides 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 provides 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, this report contains an appendix addressing Total Survey Error (TSE) and the NBS.

- **NBS Public-Use File Codebook** (Rall et al. 2012). This codebook provides extensive documentation for each variable in 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 for variables on the public-use file. The codebook also includes frequency distributions and means as appropriate.

- **NBS Questionnaire** (Wright et al. 2012). This document contains all items on the Round 4 NBS and includes documentation of skip patterns, question universe specifications, text fills, interviewer directives, and consistency and range checks.

- **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 Round 4 of the NBS. It includes an overview of the variable naming, coding, and construction conventions used in the data files and accompanying codebooks; describes how the sampling weights were computed to the final post-stratified analysis weights for both the Representative Beneficiary Sample and Ticket Participant Sample (and describes the procedures for combining the samples); outlines 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** (Barrett et al. 2012). This report describes the data processing procedures performed for Round 4 of the NBS. It outlines the data coding and cleaning procedures and describes data problems, their origins, and the corrections implemented to create the final data file. The report describes data issues by sections of the interview and concludes with a summary of types of problems encountered and general recommendations.

- **NBS Nonresponse Bias Analysis** (Grau et al. 2012). The purpose of this report is to determine if the nonresponse adjustments applied to the sampling weights of the Round 4 NBS appropriately account for differences between respondents and nonrespondents, or if the potential for nonresponse bias still exists.

The following restricted use reports are available from SSA through a formal agreement:

- **NBS Restricted-Access Codebook** (Rall et al. 2012). This codebook provide extensive documentation for each variable in 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 for variables on the restricted-access file. The codebook also includes frequency distributions and means as appropriate.

In this report, we document the editing, coding, imputation, and weighting procedures as well as the development of standard errors for the Round 4 NBS. In Chapter II, we provide an overview of the variable naming, coding, and construction conventions used in the data files and accompanying codebooks. In Chapter III, we discuss how the initial sampling weights were computed to the final post-stratified analysis weight for both the Representative Beneficiary and Ticket Participant samples and outline the procedures for combining the two samples. In Chapter IV, we describe the procedures used to impute missing responses for selected questions, and in Chapter V, we explain 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 as discussed in Chapter III are the focus of Appendix D for the cross-sectional models. In Appendix E, we present the SUDAAN and SAS parameters for the national estimates from the TTW Round 4 sample, including both the Representative Beneficiary and Ticket Participant samples, as well as the combination of the two samples.

II. DATA EDITING AND CODING

Prior to imputation, the NBS data were edited and coded to create the NBS data file. In this chapter, we document the variable naming, coding, and construction conventions used in the data files and accompanying codebooks.

A. Data Editing

At the start of data cleaning, we conducted a systematic review of the frequency counts of the individual questionnaire items. 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. As in earlier rounds, we edited only those cases with an obvious data entry or respondent error. As a result, even though we devoted considerable time to a meticulous review of individual responses, we acknowledge that some suspect values remain on the file. (See Barrett et al. (2012) 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, and dollar amounts), we reviewed the upper and lower values assigned by interviewers. While data entry ranges were set in the CATI instrument to prevent the entry of improbable responses, the ranges were set to accommodate a wide spectrum of values in order to account for the diversity expected in the population of interest and to permit the interview to continue in most situations. For these reasons, we set extremely high and low values to missing (.D = don't know) in the case of apparent data entry error.

The NBS instrument included several consistency edit checks to flag potential problems during the interview. To minimize respondent burden, however, all consistency edit checks were suppressible. While the interviewer was instructed to probe inconsistent responses, the interview could continue beyond a particular 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 such cases, we corrected them or set them to missing when we encountered an obvious error.

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 assigned numeric codes, we examined frequencies to ensure the assignment of valid values. For health condition coding, we 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 includes several 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, when possible, assigned them numeric codes. The methodology used to code each variable depended on the content of the variable.

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

Three types of questions (described below) in the NBS did not have designated response categories; rather, the responses to the questions were recorded verbatim:

- **Open-ended questions** have no response options specified (such as Item E43, Why are you no longer receiving services from your employment network?). For these items, interviewers recorded the verbatim response. Drawing on 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 the categories, it was coded as "other."
- **"Other/specify"** is a response option for questions with 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 such questions, respondents were asked to specify an answer to "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 seemed to apply, interviewers selected an "other specify" category and typed in the response.

As part of data processing in 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. We also added supplemental response categories 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, 2, and 3 coding were applied in Round 4, with no new categories added in Round 4. In some cases, categories developed in earlier rounds were added to the questionnaire to minimize back-coding. If, during coding, it became apparent that changes to the coding scheme were needed (for example, adding categories or clarifying coding decisions), we discussed and documented new decision rules. 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 previously coded cases. When it was impossible to code a response, when a response was invalid, or when a response could not be coded into a given category, we assigned a two-digit supplemental code to the response (Table II.1). The data files exclude the verbatim responses. (See Barrett et al. (2012) for full details on 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 and should be deleted
95	Refused	Used only if verbatim response indicates respondent refused to answer the question
96	Duplicate response	Indicates the verbatim response already has 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 4.

2. Health Condition Coding

Responses to questions on health conditions required a specific type of open-ended coding. Section B of the questionnaire asked each respondent to cite the main and secondary physical or mental conditions that limit the kind or amount of work or daily activities that he or she can perform. Main conditions could be reported as one of four items: Item B2 (main reason limited), Item B6 (main reason eligible for benefits), Item B12 (main reason originally eligible for benefits if not currently eligible), and Item B15 (main reason limited when first received disability benefits). The overall purpose of Items B6, B12, and B15 was to collect information on a health condition from people who reported no limiting conditions in Item 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 (Item 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 (Item B12) or for the reason that first made them eligible if they no longer had that condition (Item 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 Items B2, B6, B12, and B15 was assigned a value for the three health condition constructs. Although respondents were asked to cite one "main" condition in Items B2, B6, B12, or B15, many listed more than one. These additional responses were maintained under the main condition variable and coded in the order in which they were recorded.

For each item on a main condition, respondents were also asked to list any other, or secondary, conditions. For example, respondents reporting a main condition in Item B2 were asked in Item B4 to list other conditions that limited the kind or amount of work or daily activities they could do. Respondents reporting the main reason for their eligibility for disability benefits (Item B6) were asked in Item 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 in Item B12 (the condition that made them eligible to receive disability benefits in the last five years) were asked in Item 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 were asked for the main reason for their initial limitation were also asked if any other conditions had limited them when they started receiving benefits (Item B17).

As in previous rounds, 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 developed in 1950 to index hospital records by disease for data storage and retrieval. The ICD-9 was available in hard copy for each coder. Coders, many of whom had medical coding experience, attended an eight-hour training session before coding and were instructed to code to the highest possible level of specificity. Responses not specific enough for a five-digit code were coded to four (subcategory) or three digits (category codes). Responses 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. In Table II.2, we list the broad categorical and supplementary codes. For cases in which the respondent reported several distinct conditions, all conditions were coded (for instance, three distinct conditions would be recorded and coded as B2_1, B2_2, and B2_3).

Table II.2. ICD- 9 Category and Supplemental Codes

Code	Label	Description of ICD-9 Codes	Corresponding ICD-9 Codes
00	Other	Other and unspecified infectious and parasitic disease; alcohol dependence syndrome and drug dependence; learning disorders and developmental speech or language disorders; complications of medical care, not elsewhere classified (NEC)	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

Table II.2 *(continued)*

Code	Label	Description of ICD-9 Codes	Corresponding ICD-9 Codes
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 puerperium	Complications related to pregnancy or delivery and complications of puerperium	630-677
12	Diseases of skin/ subcutaneous tissue	Infections of the skin, inflammatory conditions, and other skin diseases	680.0-709.9
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 condition or symptom to code	No ICD-9 codes
98	Don't know	The respondent reports that he or 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 4.

We employed several means to ensure that responses were coded according to the proper protocols. We performed an initial quality assurance check, per coder, for the first several cases that were coded. In addition, during coding, 10 percent of responses were randomly selected for review. In total, a supervisor reviewed approximately 16 percent of all coded responses, including cases flagged by coders for review because coders were unable or did not know how to code them. Approximately 17 percent of all cases were recoded. In the course of the various reviews, the development of additional decision rules clarified and documented the coding protocol. The decision rules 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, with the addition of new decision rules, previously coded responses were reviewed and recoded if necessary. After completion of the ICD-9 coding, we processed the health condition variables into a series of constructed variables that grouped health conditions into broad disease groups.

3. Industry and Occupation

Information about a sample member's current employment and employment in 2009 was recorded in Section C and D of the questionnaire respectively. For each job, respondents were asked to record their occupation (Items C2 and D4) and the type of business or industry (Items C3 and D5) in which they were employed. Verbatim responses were coded using the Bureau of Labor Statistics's 2000 Standard Occupational Classification (SOC).[10] 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 groups). Appendix B lists the three-digit minor groups classified within major groups.

As for earlier rounds, verbatim responses to the industry items were coded according to the 2002 North American Industry Classification System (NAICS).[11] 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 uses 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 subsector. (Appendix C lists the broad industry sectors.) Most Federal surveys use both the SOC and NAICS coding schemes, thus providing uniformity and comparability across data sources.

Mathematica developed supplemental codes for responses to questions about occupation and industry that could not be coded to a three-digit SOC or NAICS code (Table II.3). As we did in the health condition coding, we performed an initial quality assurance check, per coder, for the first several cases coded and, during coding, randomly selected 10 percent of responses for review. In

---

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

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

total, a supervisor reviewed approximately 12 percent of all coded responses, including cases that coders flagged for review because they were unable or did not know how to code them. Approximately 15 percent of all cases required recoding.

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 or her occupation or type of business.
97	No occupation or industry reported	No valid occupation or industry is reported in the verbatim response.
98	Don't know	The respondent reports that he or she does not know the occupation or industry.
99	Uncodeable	A code cannot be assigned based on the verbatim response.

Source: NBS, Round 4.

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

III. SAMPLING WEIGHTS

The final analysis weights for the Representative Beneficiary Sample and Ticket Participant Sample were determined via a four-step process: (1) calculate the initial probability weights, (2) adjust the weights for two phases of nonresponse (location and completion), (3) trim the weights to reduce the variance, and (4) post-stratification. In Section A, we summarize the procedures used to compute and adjust the sampling weights as well as the procedure for creating composite weights.[12] In Sections B and C, respectively, we describe the procedures for computing the weights for the Representative Beneficiary Sample and the Ticket Participant Sample.

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. We selected the Representative Beneficiary Sample in two stages by (1) selecting primary sampling units (PSU) as part of the Round 1 sampling activities and (2) selecting the individuals within the PSUs from a current database of beneficiaries.[13] The Round 1 PSUs were the first-stage sampling units for all subsequent rounds. We selected 79 of these PSUs, with 2 PSUs—Los Angeles County, California, and Cook County, Illinois—acting as certainty PSUs because of their large size.[14] The Los Angeles PSU received a double allocation because it deserved two selections. The sample of all SSA beneficiaries (Representative Beneficiary Sample) was selected from among beneficiaries residing in these 79 PSUs. For the Representative Beneficiary Sample, the Los Angeles County and Cook County PSUs had a much larger number of beneficiaries than other counties, and were therefore partitioned into a large number of Secondary Sampling Units (SSUs) based on beneficiaries' ZIP codes. From these SSUs, we selected four SSUs from the Los Angeles PSU and two from the Cook County PSU.[15] Beneficiaries were selected from the PSUs or SSUs using age-defined sampling strata. In total, we selected SSA beneficiaries from 83 locations (77 PSUs and 6 SSUs) across the 50 states and the District of Columbia.

---

[12] For the Ticket Participant Sample, we combined, when needed, the supplemented stratified sample with the PSU-based Ticket Participant Sample, using a composite weight. We also combined the Representative Beneficiary Sample with the Ticket Participant Sample, using composite weights.

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

[14] Los Angeles County includes the city of Los Angeles; Cook County includes the city of Chicago.

[15] It was therefore possible for a beneficiary to reside in one of the selected PSUs (Los Angeles County or Cook County) and not be selected because they did not reside in one of the selected SSUs.

We sampled beneficiaries in the selected PSUs who were in active pay status as of June 30, 2009.[16] We used four age-based strata in each PSU. In particular, we stratified beneficiaries into the following age groups: 18- to 29-year-olds, 30- to 39-year-olds, 40- to 49-year-olds, and 50-year-olds and older. Because we used a composite size measure to select the PSUs, we could achieve equal probability samples in the age strata and nearly equal workload in each PSU for the Representative Beneficiary Sample.[17]

For the initial beneficiary sample, we selected more individuals than we expected to need in order to account for differential response and eligibility rates in both the PSUs and the sampling strata. We randomly partitioned this augmented sample into subsamples (called "waves") and used some of the waves to form the actual final sample (i.e., released for data collection). We released an initial set of waves and then monitored data collection to identify which PSUs and strata required additional sample members. After we released sample members in the initial waves, we were able to limit the number of additional sample members (in subsequently released waves) just to those PSUs and strata requiring them and thus achieved sample sizes close to our targets while using the smallest number of beneficiaries. Controlling the release of the sample also allowed us to control the balance between data collection costs and response rates. We computed the initial sampling weights based the inverse of the selection probability for the augmented sample. Given that we released only a subset of the augmented sample, we then adjusted the initial weights for the actual sample size. The release-adjusted weights were post-stratified to population totals obtained from SSA.[18]

We then needed to adjust the initial sampling weights for nonresponse. A commonly used method for computing weight adjustments is to form classes of sample members with similar characteristics and then 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. We formed the "weighting classes" in such a way to ensure that there were sufficient counts in each class to make the adjustment more stable (that is, to ensure smaller variance). The natural extension to the weighting class procedure is to perform logistic regression with the weighting class definitions used as covariates, provided that each level of the model covariates has a sufficient number of sample members to ensure a stable adjustment. The inverse of the propensity score is then the

[16] Beneficiaries with selected non-payment status codes were included only if the denial variable was blank. However, based on our experience in prior rounds, we received an updated data extraction after sampling and prior to fielding to identify beneficiaries who may have been in a "holding" status at the time of sample selection, but who had subsequently been denied benefits. These cases were coded as ineligible prior to fielding. Due to time constraints, this extraction was limited to SSI files at Round 4. Hence, the payment-type distribution among ineligible cases contains more SSI-only cases and fewer SSDI-only cases than would be expected if the ineligible cases were like the rest of the population. We also statused as ineligible any beneficiaries who died between June 30, 2009 and the start of data collection based on information obtained from LexisNexis\Accurint prior to the start of data collection. Additionally, beneficiaries who were found to be deceased, incarcerated, no longer living in the continental United States, or reported that they had not received benefits in the past five years at the time of the interview, were statused as ineligible during the data collection period. The proportion of cases statused as ineligible during data collection (6.0%) was similar to the ineligibility rates obtained during prior rounds (6.4% in round 3, 5.6% in round 2, and 5.1% in round 1) and the impact on yield rates was negligible.

[17] 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).

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

adjustment factor. 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 nonresponse weight adjustments (at both the location and cooperation stages), we used logistic models to estimate the propensity for a sample member to respond. The adjusted weight for each sample case is the product of the initial sampling weight and the adjustment factor.

We calculated the adjustment factor in two stages: (1) by estimating a propensity score for locating a sample member and (2) by estimating a propensity score for response among these located sample members. In our experience with the NBS, factors associated with the inability to locate a person tend to differ from factors associated with cooperation. The unlocated person generally does not deliberately avoid or otherwise refuse to cooperate. For instance, that person may have chosen not to list his or her phone number or may frequently move from one address to another, even though there is no evidence to suggest that, once located, he or she would show a specific unwillingness to cooperate with the survey. Located nonrespondents, on the other hand, may deliberately avoid the interviewer or express displeasure or hostility toward surveys in general or SSA in particular.

To develop the logistic propensity models for Round 4, we used as covariates information from the SSA data files as well as geographic information (such as urban or rural region). We obtained much of the geographic information from the Area Resource File (ARF 2009–2010), a file with county-level information about the population, health, and economic-related matters for every county in the United States. Using a liberal level of statistical significance (0.3) in forward and backward stepwise logistic regression models, we made an initial attempt to reduce the pool of covariates and interactions. We used a higher significance level because each model's purpose 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. We excluded from the pool any covariate or interaction that was clearly unrelated to locating the respondent or to response propensity. Given that the stepwise logistic regression analysis does not fully account for the complex survey design, we developed the final weighted models by using SUDAAN software, which accounts appropriately for the complex sample design.

The next step called for the careful evaluation of a series of models by comparing the following measures of predictive ability and goodness of fit: the R-squared statistic, Akaike's Information Criterion (AIC),[19] the percentage of concordant and discordant pairs, and the Hosmer-Lemeshow goodness-of-fit test. 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 manipulated the set of variables to avoid data warnings in SUDAAN.[20] We then used

[19] Akaike's Information Criterion is defined as AIC = $-2\text{LogL} + 2(k+s)$, where LogL is the log likelihood of the binomial distribution using the parameters from the given model, k is the total number of response levels minus 1, 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 preferable to larger values.

[20] 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, although avoidance of the first two was of highest priority. The warnings usually were caused by a response cell with a count that was too small, which required dropping covariates or collapsing categories in covariates.

the specific covariate values for each located person to estimate the propensity score, from which the adjustment factor was determined by taking the inverse. When computing the adjustment factors, we limited the maximum location adjustment to smaller than two and the maximum cooperation adjustment to smaller than three. If such a location adjustment was not possible, we used a trimming algorithm to reallocate the part of location adjustments greater than two (or the part of the cooperation adjustments greater than three) to other individuals with smaller adjustment factors.[21] The location-adjusted weight is the product of the released-adjusted probability weight and the trimmed location adjustment. The nonresponse-adjusted weight is the product of the location-adjusted weight and the inverse of the cooperation propensity score, calculated in the same manner as the location propensity score.

Once we made the adjustments, we assessed the distribution of the adjusted weights for unusually high values, which could make the survey estimates less precise. We used the design effect attributed to the variation in the sampling weights as a statistical measure to determine both the necessity and amount of trimming. The design effect attributed 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, we checked the design effect attributable to unequal weighting within the age-related sampling strata and determined that no further trimming of the adjusted weights was required. The maximum design effect among all age strata in the Representative Beneficiary Sample was 1.10.

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, for beneficiaries only, recipient status[22]). After post-stratification, we checked the survey weights again to determine whether more trimming was needed. In Round 4, trimming was not needed after post-stratification in the Representative Beneficiary Sample.

2. Ticket Participant Sample

We computed the initial sampling weights for the Ticket Participant Sample on the basis of the inverse of the selection probability for the participant. As with the Representative Beneficiary Sample, we used the PSUs as the primary source of sample members and, when possible, selected an initially larger (augmented) sample. We sampled participants from the selected PSUs for the clustered sample, and throughout the 50 states and the District of Columbia for the unclustered sample. We selected participants who had used a ticket at least once on January 1, 2009, or between January 1, 2009, and October 2, 2009.[23] We selected the sample of all TTW participants (Ticket

---

[21] This is a form of weight trimming. Among the location adjustments, 26 cases were trimmed, and 5 cases had a trimmed cooperation adjustment factor (discussed in Section B.2.d of this chapter).

[22] Disability payments were made in the form of SSI or SSDI or both.

[23] Individuals in the TTW Participant sample were not evaluated based on pay status since they were determined to be "Ticket eligible" by SSA. As for the Representative Beneficiary sample, we statused as ineligible any participants who died between October 2, 2009 and the start of data collection based on information obtained from LexisNexis\Accurint prior to the start of data collection. Additionally, participants who were found to be deceased, incarcerated, no longer living in the continental United States, or reported had not received benefits in the past five years at the time of the interview, were statused as ineligible during the data collection period. The proportion of cases statused as ineligible during data collection (1.8%) was similar to the ineligibility rates obtained during prior rounds (1.1% in round 3, 1.5% in round 2, and 2.7% in round 1) and the impact on yield rates was negligible.

Participant Sample) from among participants residing in the same PSUs and used no secondary sampling units.[24] In all four rounds of the NBS, the number of Ticket participants in the selected PSUs was insufficient in one or more participant strata for the analysis. For such strata, we drew a supplemental single-stage sample from all Ticket participants, those both in and not in the PSUs, with stratification based on payment type (Rounds 1 through 3) or provider and payment type (Round 4) and whether the participant was or was not in a PSU.

For participants with Tickets assigned either to SVRAs acting as ENs or non–SVRA ENs, the PSUs in the initial sampling design lacked a sufficient number of participants to support the analysis tasks—even with all participants in the PSUs from these two provider-payment types selected for the sample. As a result, we had to supplement the sample from the PSUs with a second independent sample of Ticket participants from two geographic strata defined by the PSUs (participants residing in a PSU or not residing in any of the PSUs).[25] We refer to the initial sample design as the "clustered" sample; the second independent sample is referred to as the "unclustered" sample. Mathematica randomly selected sample members in the unclustered sample in the two aforementioned geographic strata from the entire population of participants with Tickets assigned to SVRAs receiving traditional CR payments and participants with tickets assigned to non–SVRA ENs.[26] We referred to the combination of data from the clustered and unclustered samples to calculate estimates as a paired sample design (discussed later).

As with the Representative Beneficiary Sample, we computed the weights for the augmented sample and then adjusted them for the number of sample members released into the final sample. We adjusted for located sample members and then for response among such members. We used logistic propensity models to calculate the location adjustment for all Ticket participants and the response adjustments for located Ticket participants of all three provider-payment types. As needed, we trimmed adjustments so that they did not exceed two for the location model and three for the cooperation model.[27] The modeling procedures were similar to those used with the Representative Beneficiary Sample.

The size of the sample for the three provider-payment types was similar, but the size of the population for each was markedly different. (More than 80 percent of the population of Ticket participants had their Ticket assigned to an SVRA under the traditional payment system. In Section C, we provide percentages for each phase and provider-payment type.) Hence, the sampling weights differed substantially in magnitude from one provider-payment type to the next. As a result, we conducted the weight adjustments separately for each provider-payment type. For the subsamples associated with provider-payment type within the Ticket Participant Sample, we trimmed the

---

[24] For the Ticket Participant Sample, Mathematica selected participants from the entire Los Angeles County PSU and from the entire Cook County PSU.

[25] Given that the target population for the NBS did not include Puerto Rico or other outlying territories, we excluded from the frame all beneficiaries and Ticket participants who resided in these areas.

[26] Because of the small populations for the provider types where the paired sample design was required, Mathematica often selected Ticket participants who resided in the selected PSUs for these provider types for both the clustered and in-PSU strata of the unclustered samples. Hence, we had to count these duplicate cases in the weighting process (discussed later).

[27] Across the three Ticket participant subpopulations, we trimmed 11 location adjustment factors and 4 cooperation adjustment factors (details in Section C.2.d of this chapter).

weights to ensure that the design effect attributable to unequal weighting was not substantially greater than 3.0 (less than 3.0, if possible). (In Section C, we provide more detail on the trimming of participants' weights and the design effects attributable to unequal weighting before and after trimming.) The final adjustment for 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 the need for more trimming. Even though the Round 4 weights required trimming before post-stratification in the Ticket Participant Sample, they required no further trimming after post-stratification.

3. Composite Weights

While the Ticket participant population constitutes a small subset of the beneficiary population, some analyses required a sample with a substantial number of individuals both within and outside the Ticket participant population. Such a sample simply represents a combination of the Ticket Participant and Representative Beneficiary samples and required the use of composite weights to account for the combined sample. When conducting analyses representing the beneficiary population, we used the combined sample weights to make estimates about participants within the beneficiary population. (Analyses limited to the participants' subpopulation used weights from the Ticket Participant Sample only.)

In Round 1, we used a sophisticated procedure to create the weights in order to minimize the variance of survey estimates. The procedure allowed weights to be applied to observations duplicated across the two samples.[28] However, given that Ticket participants were such a small fraction of the beneficiary sample frame, we used a simpler alternative method in Rounds 2, 3, and 4.

In Round 4, we replaced the original Representative Beneficiary Sample weights with a value of zero among the 50 Ticket participants selected for that sample. To ensure representation of the Ticket participant population, we replaced these 50 members of the Representative Beneficiary Sample with the 4,334 members of the Ticket Participant Sample who had completed an interview (or had ineligible dispositions after sample selection). The sum of the weights for the 50 participants in the Representative Beneficiary Sample is an unbiased estimate of the number of participants in the sampling frame. However, given the relatively small number of Ticket participants in the Representative Beneficiary Sample, the estimate did not equal the known total in the sampling frame, as expected. The post-stratification adjustment realigned the population totals.

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 the Representative Beneficiary and Ticket Participant cross-sectional samples as well as the composite weights. For the sake of objectivity, we chose a statistician not directly involved in the project.

[28] A complex procedure also combined the clustered and unclustered samples of the Ticket Participant Sample in all rounds (described in Section C of this chapter).

B. Representative Beneficiary Sample

1. Initial Weights

We computed the initial weights by using the inverse of the probability of selection. For the Representative Beneficiary Sample, we selected samples independently in each of four age strata in each geographic unit or PSU.[29] We determined the number of sample members selected in each stratum and PSU for the augmented sample by independently allocating five times the target sample size across the 83 geographic units (PSUs and secondary sampling units) for each stratum,[30] thereby ensuring the availability of ample reserve sample units in case response or eligibility rates were lower than expected. The augmented sample size for the three younger age strata (18- to 29-year-olds, 30- to 39-year-olds, and 40- to 49-year-olds) was 3,335 sample members (roughly five times the target sample size of 667); for beneficiaries age 50 and older, the augmented sample size was 1,998 (again, five times the target sample size of 400). By using the composite size measure already described, we calculated the initial weights for the full augmented sample of 12,000 sample members by taking the inverse of the global sampling rate (F_j) for each stratum. In Table III.1, we provide the global sampling rates and initial weights.

Table III.1. Survey Population as of June 30, 2009, Initial Augmented Sample Sizes and Initial Weights by Sampling Strata in the National Beneficiary Survey

Sampling Strata (ages as of June 30, 2009)	Survey Population[a]	Augmented Sample Size	Global Sampling Rate (Fj)	Initial Sample Weights	Released Sample
Beneficiaries age 18 to 29	1,295,767	3,335	0.002574	388.5	1,029
Beneficiaries age 30 to 39	1,314,526	3,335	0.002537	394.2	1,032
Beneficiaries age 40 to 49	2,524,579	3,335	0.001321	757.00	1,019
Beneficiaries age 50 and older	6,982,459	1,998	0.000286	3,496.5	603
Total	12,117,331	11,999			3683

Source: Sample allocation and counts computed by Mathematica.

[a] The survey population represents all SSI and SSDI beneficiaries in the 50 states and the District of Columbia. It excludes 185,840 beneficiaries who live in Puerto Rico and other outlying territories.

As described previously, we randomly partitioned the full sample 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 83 geographic units (a total of 332 combinations of PSUs and sampling strata). At the start of data collection, we assigned a preliminary sample to the data collection effort and then assigned additional waves as needed, based on experience with eligibility and response rates. Within the 336 combinations of PSUs and sampling strata, we adjusted the initial weights to

[29] The sample of PSUs contained 79 unique selections. Given the size of its beneficiary population, the PSU representing Los Angeles County (LA) received two selections. Within the LA PSU, we formed SSUs and selected four. In the PSU representing Cook County (Chicago), we also formed SSUs in order to decrease travel costs and selected two. The six SSUs and the other 77 PSUs (83 units) were treated as PSUs for the beneficiary sample.

[30] We selected an augmented sample that was five times as large as needed in order to allow for both an 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.

account for the number of waves assigned to data collection. The final sample size for the Representative Beneficiary Sample totaled 3,683 beneficiaries, as shown under "Released Sample" in Table III.1.

2. Nonresponse Adjustment

As in virtually all surveys, we had to adjust the sampling weights to compensate for sample members who could not be located or who, once located, refuse to respond. First, we fitted weighted logistic regression models where the binary response was whether the sample member could be located. Using variables obtained from SSA databases, we selected, through stepwise regression, a pool of covariates from which to choose a final location model. The pool included both main effects and interactions. From the pool of covariates, we used various measures of goodness of fit and predictive ability to compare candidate models while avoiding large adjustments. Even though we developed the logistic regression propensity models to minimize the number of large adjustment factors, we still had to trim the adjustment factors, within trimming classes based on the four age categories in order to ensure that the maximum did not exceed two. We repeated the process for interview respondents among the located sample members and fitted another weighted logistic regression model, trimming large adjustments within the four age categories so that the maximum did not exceed three.[31] The two levels in the binary response for this model were "respondent" or "nonrespondent." For the Representative Beneficiary Sample, a sample member was classified as a respondent if the sample member or the person responding for the sample member completed the interview (that is, an eligible respondent) or if the sample member was deemed ineligible after sample selection (an ineligible respondent). Ineligible sample members included persons who were never SSA beneficiaries, were in the military at the time of the survey, were incarcerated, had moved outside the United States, or were deceased at the time of the survey.

Based on the above procedures, the main factors or attributes affecting our ability to locate and interview a sample member included the sample member's personal characteristics (race, ethnicity, gender, and age), identity of the payee with respect to the beneficiary, whether the beneficiary and the applicant for benefits lived in the same location, how many phone numbers or addresses were n the SSA files for the beneficiary, living situation of beneficiary, and geographic characteristics, including attributes of the county where the beneficiary lives.

a. Coding of Survey Dispositions

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

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

[31] As stated earlier, we trimmed 26 location adjustment factors and 5 cooperation adjustment factors (discussed in Section B.2.d of this chapter).

3. Located nonrespondents (including active or passive refusals and language barrier situations).
4. Unlocated sample members (sample members who could not be located through either 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.[32]

b. Response Rates

The 72.8 percent **response rate** for the Representative Beneficiary Sample noted in the introduction to this report and given in the first line of Table III.2, is the weighted count of sample members who completed an interview or were deemed ineligible, divided by the weighted sample count of all sample members.[33] It may 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.

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 93.3 percent. The **weighted cooperation rate** (the weighted completion rate among located sample members), 77.8 percent in Table III.2, is the weighted count of sample members who completed an interview or were deemed ineligible, divided by the weighted sample count of all located sample members. Weighted cooperation rates reflect the common survey situation that, once a person is located, repeated contact efforts often result in a completed interview.

---

32 Disposition codes 420 (institutionalized) and 430 (unavailable during field period) were classified as nonrespondent codes in Rounds 2, 3, and 4, even though they were considered ineligible codes in Round 1. This affected cases in the beneficiary samples of Round 2 (eight cases), Round 3 (six cases), and Round 4 (five cases). As a result, the nonresponse adjusted weight for these cases was zero in Rounds 2, 3, and 4, even though a similar response in Round 1 would have resulted in a positive weight. In view of the small numbers, the effect on response rates was very small.

33 The response rate is calculated as the weighted count of sample members who completed an interview or were deemed ineligible divided by the weighted sample count of all sample members: (number of completed interviews + number of partially completed interviews + number of ineligibles)/(number of cases in the sample). The response rate is essentially equivalent to the American Association of Public Opinion Research (AAPOR) standard response rate calculation, assuming that all nonrespondents have unknown eligibility status: RR_{AAPOR} = number of completed interviews/(number of cases in the sample - estimated number of ineligible cases). Ineligible cases are included in the numerator and denominator for two reasons: (1) the cases classified as ineligible are part of the original sampling frame (and hence the study population), and we obtained complete information for fully classifying these cases (that is, their responses to the eligibility questions in the questionnaire are complete) such that we may classify them as respondents; and (2) incorporation of the ineligibles into the numerator and denominator of the response rate is essentially equivalent to the definition of a more conventional response rate, assuming that all nonrespondents have unknown eligibility status.

Table III.2. Weighted Location and 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	3,683	3,345	93.3	2,520	77.8	72.8
SSI Only, SSDI Only, or Both SSI and SSDI						
SSI only	1,581	1,404	90.4	1,068	78.7	71.6
SSDI only	1,322	1,237	96.1	922	76.8	74.0
Both SSI and SSDI	780	704	90.4	530	79.1	71.8
SSI or SSDI						
SSI only or both SSI and SSDI	2,361	2,108	90.4	1,598	78.9	71.7
SSDI only or both SSI and SSDI	2,102	1,941	94.6	1,452	77.4	73.4
Constructed Disability Status						
Deaf	44	40	92.5	29	80.1	75.5
Mental	2,016	1,811	91.8	1,333	76.7	70.6
Physical	1,488	1,379	94.6	1,071	78.7	74.8
Unknown	135	115	89.1	87	76.4	68.0
Beneficiary's Age (four categories)						
18 to 29	1,029	921	90.0	705	77.7	70.2
30 to 39	1,032	912	88.5	679	75.4	67.3
40 to 49	1,019	941	92.7	699	75.7	70.5
50 and older	603	571	95.0	437	79.1	75.2
Sex						
Male	1,935	1,751	93.3	1,297	76.4	71.5
Female	1,748	1,594	93.2	1,223	79.2	74.2
Hispanicity						
Hispanic	250	214	89.3	153	63.9	58.9
Non-Hispanic	3,433	3,131	93.5	2,367	78.6	73.6
Race						
White	2,115	1,955	94.1	1,465	77.1	72.7
Black	857	761	92.4	594	82.7	76.6
Unknown	628	555	92.2	413	75.7	70.8
Asian American, Pacific Islander,	57	51	85.0	27	36.8	29.8
North American Indian, or Alaskan Native	26	23	78.8	21	94.0	74.4
Living Situation						
Living alone	2,362	2,124	91.8	1,622	79.3	73.2
Living with others	273	252	88.9	187	80.8	72.2
Living with parents	72	60	84.4	40	68.3	57.6
In institution or unknown	976	909	96.0	671	75.6	72.7
Did the Applicant for Benefits Live in Same ZIP Code as the Beneficiary?						
No	387	331	84.8	236	72.9	62.6
Yes	2,199	1,998	93.5	1,538	80.6	75.7
No information	1,097	1,016	94.8	746	75.0	71.0

Table III.2 *(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	108	99	95.6	65	60.1	58.4
Payee is a family member	1,262	1,138	90.5	856	76.9	69.9
Payee is an institution	220	200	90.1	150	77.2	69.9
Other	2,093	1,908	94.1	1,449	78.7	74.2
Count of Phone Numbers in File						
Only one phone number in file	59	58	98.8	48	82.4	81.2
Two phone numbers in file	513	491	96.1	394	80.9	77.9
Three phone numbers in file	331	294	92.2	210	74.7	69.5
Four phone numbers in file	265	228	85.9	168	77.6	68.4
Five or more phone numbers in file	1,056	851	83.2	623	76.9	63.9
No information	1,459	1,423	98.7	1,077	77.8	76.8
Count of Addresses in File						
One address in file	1,416	1,336	96.6	1,065	81.2	78.5
Two addresses in file	1,017	872	88.5	628	73.9	65.2
Three or more addresses in file	417	313	73.8	207	71.0	52.5
No information	833	824	99.2	620	78.8	78.2
Type of Claim						
Survivor	393	369	92.7	261	74.6	69.3
Disabled	1,787	1,644	95.0	1,244	78.0	74.3
Unknown	1,503	1,332	90.2	1,015	78.4	71.1
Census Region						
Midwest	888	798	92.6	604	78.1	72.4
Northeast	583	532	94.0	389	77.7	73.4
South	1,501	1,387	95.0	1,072	79.5	75.8
West	711	628	89.5	455	73.6	66.0
Census Division						
East North Central	630	559	91.7	432	81.7	75.1
East South Central	286	269	94.7	215	83.5	79.3
Middle Atlantic	400	364	94.1	267	79.8	75.3
Mountain	190	170	90.1	128	76.6	68.7
New England	183	168	93.9	122	73.0	69.1
Pacific	521	458	89.3	327	72.5	65.0
South Atlantic	839	767	94.5	572	76.7	72.7
West North Central	258	239	94.9	172	68.6	65.4
West South Central	376	351	96.2	285	82.2	79.5
Metropolitan						
Metropolitan areas of 1 million population or more	1,592	1,438	92.7	1,062	76.9	71.4
Metropolitan areas of 250,000 to 999,999 population	962	870	93.2	649	78.2	73.4
Metropolitan areas of fewer than 250,000 population	409	368	92.7	273	74.1	68.6
Nonmetropolitan areas adjacent to large metropolitan areas	267	250	95.8	199	77.6	74.6
Nonmetropolitan areas adjacent to medium or small metropolitan areas	257	242	96.5	199	86.8	83.9

Table III.2 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Nonmetropolitan areas not adjacent to metropolitan areas	196	177	91.1	138	76.4	69.7
County with Low Education						
Yes	565	518	94.9	394	76.7	73.3
No	3,118	2,827	93.0	2,126	78.0	72.7
County with Housing Stress						
Yes	1,535	1,384	92.4	1,022	76.2	70.8
No	2,148	1,961	93.8	1,498	78.9	74.2
Population Loss County						
Yes	395	361	94.9	283	84.4	80.6
No	3,288	2,984	93.1	2,237	77.0	71.9
Retirement Destination County						
Yes	498	443	90.5	343	77.6	70.4
No	3,185	2,902	93.7	2,177	77.8	73.2
Service- Dependent Economy County						
Yes	1,556	1,403	91.4	1,018	74.9	68.8
No	2,127	1,942	94.6	1,502	79.8	75.6
Nonspecialized- Dependent Economy County						
Yes	1,023	922	94.5	719	79.8	75.5
No	2,660	2,423	92.8	1,801	77.1	71.8
Government- Dependent Economy County						
Yes	349	322	93.7	234	73.8	69.2
No	3,334	3,023	93.2	2,286	78.2	73.2
County Racial/Ethnic Profile						
County with at least 90% non-Hispanic white population	594	545	94.8	435	79.5	75.4
County with plurality or majority Hispanic population	392	348	93.0	263	74.4	69.8
County with majority but fewer than 90% non-Hispanic white population	1,413	1,281	93.1	953	77.8	72.8
County with a racially/ethnically mixed population, no majority group	1,089	989	92.3	712	75.7	69.9
County with plurality or majority non-Hispanic black population	142	132	97.2	114	91.0	88.9
County with at least 20% American Indian population	53	50	91.2	43	84.7	77.2
Phase						
Phase 1	1,086	981	92.9	716	75.8	70.7
Phase 2	1,124	1,022	93.2	785	79.2	73.9
Phase 3	1,473	1,342	93.6	1,019	78.2	73.5

Source: NBS, Round 4.

We use the weighted rates 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 for use in the data analysis or in determining ineligibility for the analysis.

c. Factors Related to Location and Response

In addition to overall response rate information, Table III.2 provides information for selected factors associated with locating a sample member and for factors associated with response among located sample members. The table displays the unweighted counts of all sample members, counts of located sample members, and counts of sample members who completed an interview or were deemed ineligible. We also include in the table 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 to be used in the nonresponse adjustment models.

d. Propensity Models for Weight Adjustments

Using the main effects already described as well as selected interactions, we developed response propensity models to determine the nonresponse adjustments. To identify candidate interactions from the main effects for the modeling, we first ran a chi-squared automatic interaction detector (CHAID) analysis in SPSS to find possible significant interactions.[34] 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 showing all statistically significant subgroups. The algorithm identifies splits in the population, which are as different as possible based on a chi-square statistic. The forward stepwise procedure finds the most diverse subgroupings and then splits each subgroup further into more diverse sub-subgroups. Sample size limitations are set to avoid cells with small counts. The procedure stops when splits are no longer significant; that is, a group is homogeneous with respect to variables not yet used, or cells contain too few cases. The CHAID procedure produces a tree that identifies the set of variables and interactions among the variables that are associated with the ability to locate a sample member (and a located sample member's propensity either to respond to or to be deemed ineligible for the NBS). We first ran CHAID with all covariates and then re-ran it a few times with the top variable in the tree removed to ensure the retention of all potentially important interactions for additional consideration. We further reduced the resulting pool of covariates 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, the levels were

[34] CHAID is normally attributed to Kass (1980) and Biggs et al. (1991), and its application in SPSS is described in Magidson (1993).

collapsed if collapsing was possible. If collapsing was not possible, then we excluded the covariate or interaction from the pool.[35]

To refine further the candidate variables and interaction terms, we then processed all of the resulting candidate main effects and the interactions identified by CHAID using forward and backward stepwise regression (using the STEPWISE option of the SAS LOGISTIC procedure with weights normalized to the sample size).[36] After identifying a smaller pool of main effects and interactions for potential inclusion in the final model, we carefully evaluated a set of models to determine the final model. Given that the SAS logistic regression procedure does not incorporate the sampling design, we relied on the logistic regression procedure in SUDAAN to make the final selection of covariates.

For selecting variables or interactions in the stepwise procedures, we included variables or interactions with a statistical significance level (alpha level) of 0.30 or lower (instead of the commonly used 0.05).[37] Once we determined the candidate list of main effects and interactions, we used a thorough model-fitting process to determine a parsimonious model with few very small propensities. (In Section A of this chapter, we described the model selection criteria.) In Table III.3, we summarize the variables used in the model as main effects and interactions for locating a sample member and, in Table III.4, for cooperation among located sample members.

Table III.3. Location Logistic Propensity Model: Representative Beneficiary Sample

Factors in Location Model
Main Effects
MOVE (COUNT OF ADDRESSES IN SSA FILES)
PHONE (COUNT OF PHONE NUMBERS IN SSA FILES)
METRO (METROPOLITAN STATUS OF COUNTY)
REGION (CENSUS REGION)
RACE
CNTYRACE (COUNTY RACIAL/ETHNIC PROFILE)
CNTYPOPLOSS (POPULATION LOSS COUNTY)
CNTYLOWEDUC (LOW EDUCATION COUNTY)
Two- Factor Interactions
RACE*CNTYRACE

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

[36] SUDAAN offers no automated stepwise procedures; the stepwise procedures described here were performed by using SAS.

[37] As stated, we used a higher significance level because the model's purpose was to improve the estimation of the propensity score rather than 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.

Table III.4. Cooperation Logistic Propensity Model: Representative Beneficiary Sample

Factors in Cooperation Model
Main Effects
AGECAT (AGE CATEGORY)
RACE
HISPANICITY
METRO (METROPOLITAN STATUS OF COUNTY)
DIVISION (CENSUS DIVISION)
GENDER (SEX)
REPREPAYEE (IDENTITY OF PAYEE WITH RESPECT TO BENEFICIARY)
PDZIPSAME (WHETHER APPLICANT FOR BENEFITS LIVES IN SAME ZIP CODE AS BENEFICIARY)
MOVE (COUNT OF ADDRESSES IN SSA FILES)
PHONE (COUNT OF PHONE NUMBERS IN SSA FILES)
LIVING (LIVING SITUATION)
CNTYRACE (COUNTY RACIAL/ETHNIC PROFILE)
CNTYGOV (GOVERNMENT-DEPENDENT ECONOMY COUNTY)
Two- Factor Interactions
PDZIPSAME*PHONE
PDZIPSAME*METRO
GENDER*METRO
HISPANICITY*AGECAT
HISPANICITY*MOVE

The R-squared is 0.107 (0.275 when rescaled to have a maximum of 1) for the location model and 0.070 (0.107 when rescaled) for the cooperation model.[38] These values are similar to those observed for other response propensity modeling efforts that use logistic regression with design-based sampling weights. For the location model, 77.9 percent of pairs are concordant, 21.2 percent of pairs are discordant,[39] and the p-value for the chi-square statistic from the Hosmer-Lemeshow (H-L) goodness-of-fit test is 0.441[40]; these values indicate a reasonably good fit of the model to the data. The location adjustment from the model, calculated as the inverse of the location propensity score, ranged from 1.00 to 2.36. To ensure that the maximum did not exceed 2.0, we trimmed 26 adjustment factors so that the location adjustment ranged from 1.00 to 2.00. For the cooperation model, 59.5 percent of pairs are concordant, 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 0.480 for the model. The cooperation adjustment from the model ranged from 1.02 to 3.78. To ensure that the maximum did not exceed 3.0, we trimmed five adjustment factors so that the cooperation adjustment ranged from 1.02 to 3.00. The overall nonresponse adjustment (the product of the location adjustment and the cooperation adjustment) ranged from 1.05 to 5.45.[41]

---

[38] 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. The "Max Rescaled R-Square" scales this value to have a maximum of 1.

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

[40] The Hosmer-Lemeshow Goodness-of-Fit Test is a test for goodness of fit of logistic regression models. Unlike the Pearson and deviance goodness-of-fit tests, it may be used to test goodness of fit even when some covariates are continuous (Hosmer and Lemeshow 1989).

[41] Recognizing that the Akaike's Information Criterion (AIC) is a relative number and has no meaning on its own, we do not provide values for it here.

Among the variables used in the location and cooperation models shown in Tables III.3 and III.4, the number of levels used in the models is often fewer than the number of levels in Table III.2; the levels collapsed for the models are described following the tables. The factors used in the location model include:

1. **MOVE.** Count of addresses in SSA files; four levels: (0) no information, (1) one address in file, (2) two addresses in file, (3) three or more addresses in file.
2. **PHONE.** Count of phone numbers in SSA files; three levels: (0) no information, (1) one phone number in file, (2) two or more phone numbers in file.
3. **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 of 250,000 to 999,999, (3) beneficiary lived in metropolitan area with population of fewer 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 fewer than 1 million, (6) beneficiary lived in nonmetropolitan area not adjacent to metropolitan area.
4. **REGION.** Geographic region (based on U.S. Census divisions) of beneficiary's place of residence; two levels: (1) South, (2) all other regions.
5. **RACE.** Race; two levels: (1) non–Hispanic white, (2) not white or not known to be white.
6. **CNTYRACE.** County racial ethnic profile; two levels: (1) county with racially/ethnically mixed population based on 2000 Census, no majority group; (2) other racial/ethnic profile in county.
7. **CNTYPOPLOSS.** County with population loss; two levels: (1) county with population loss in both 1980–1990 and 1990–2000 decennial periods, (2) county with population gain in 1980–1990 and/or 1990–2000 decennial periods.
8. **CNTYLOWEDUC.** County with low education; two levels: (1) county where 25 percent or more of residents age 25 through 64 had neither a high school diploma nor Graduate Equivalency Degree (GED) in 2000, (2) county without this attribute.

The model also included various interactions among these variables for locating sample members. In Table III.3, we provide the main effects using the variable names listed above as well as interactions. In Appendix D, we provide an expanded form of Table III.3 showing the levels of interactions shown in Table III.3 along with parameter estimates and their standard errors. The factors used in the cooperation model include:

1. **AGECAT.** Beneficiary's age category; three levels: (1) age 18 to 29, (2) age 30 to 39, (3) age 40 to 64.
2. **RACE.** Race of beneficiary; two levels: (1) non–Hispanic black, (2) not non–Hispanic black or not known to be non–Hispanic black.
3. **HISPANICITY.** Whether the beneficiary was Hispanic or not; two levels: (1) Hispanic, (2) not Hispanic or unknown.

4. **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 fewer 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 fewer than 1 million, (6) beneficiary lived in nonmetropolitan area not adjacent to metropolitan area.

5. **DIVISION.** Geographic region (based on U.S. Census divisions) of beneficiary's place of residence; three levels: (1) New England, (2) West North Central), (3) all other Census divisions.

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

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

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

9. .**MOVE.** Count of addresses in SSA files; four levels: (0) no information, (1) one address in file, (2) two addresses in file, (3) three or more addresses in file.

10. **PHONE.** Count of phone numbers in SSA files; three levels: (0) no information, (1) one phone number in file, (2) two or more phone numbers in file.

11. **LIVING.** Beneficiary's living situation: two levels: (1) beneficiary lives in institution, (2) beneficiary lives alone, with others, with parents, or living situation unknown.

12. **CNTYRACE.** County racial ethnic profile; two levels: (1) county with racially/ethnically mixed population based on 2000 Census, no majority group; (2) other racial/ethnic profile in county.

13. **CNTYGOV.** County with government-dependent economy; two levels: (1) 15 percent or more of average annual labor and proprietors' earnings derived from Federal and state government during 1998–2000, (2) county without this attribute.

Once again, we included various interactions among these variables in the model for the cooperation of sample members. In Table III.4, we provide the main effects using the variable names as well as interactions. In Appendix D, we provide an expanded form of Table III.4, with the levels of the interactions shown in Table III.4, along with parameter estimates and their standard errors.

After we applied adjustments to the sampling weights, we reviewed the distribution of weights to determine the need for further trimming of the weights. We concluded that no additional trimming was needed and that the maximum design effect attributable to unequal weighting was 1.10, observed with the third-oldest age group stratum.

3. Post-Stratification

Post-stratification is the procedure that aligns the weighted sums of the response-adjusted weights to known totals external to the survey. The process offers face-validity for reporting population counts and has some statistical benefits. For the Representative Beneficiary Sample, we post-stratified to the 24 population totals obtained from SSA.[42] 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). We conducted no trimming after post-stratification.

C. Ticket Participant Sample

As noted earlier, we selected the Ticket Participant Sample from the Round 4 population of Ticket-to-Work participants, a subset of all SSI/SSDI beneficiaries, and partitioned the sample according to the provider-payment types in the Ticket-To-Work payment system (traditional SVRAs, SVRA ENs, and non–SVRA ENs). Participants with Tickets assigned to an SVRA receiving traditional CR payments accounted for 81 percent (68,592 of 85,038) of participants at the time of sampling frame development. The number of participants with Tickets assigned to SVRAs functioning as ENs under TTW totaled 12,728 (15 percent). The number of participants with Tickets assigned to non–SVRA ENs totaled only 3,718 (4 percent).[43] As also noted earlier, the PSUs in the initial sampling design did not contain a sufficient number of participants with Tickets assigned to non–SVRSA ENS and SVRAs functioning as ENs to support analysis tasks. As a result, we supplemented the clustered sample, which consisted of respondents selected within the initial sample design, by randomly selecting a sample from the entire population of participants with Tickets assigned to ENs (non–SVRA ENs and SVRAs functioning).

Given that the clustered sample was part of the original sample design, we selected all respondents in the clustered sample from PSUs, whereas the unclustered sample included units that may or may not have been in the selected PSUs. We therefore organized the unclustered sample into two strata: in the PSU or not in the PSU. In most cases, respondents selected for the in-PSU stratum of the unclustered sample were also in the clustered sample. The weights for such duplicate cases had to be adjusted appropriately to account for a single respondent's appearance in two independent samples. (In the next subsection, we discuss the compositing scheme used to make the needed adjustments.) In addition, if the central office[44] could not locate sample members based on sample frame information, it treated them differently in the clustered and unclustered samples. For the clustered sample, the central office sent sample cases that they could not locate by telephone to the field for further follow-up for attempted personal interviews. In the unclustered sample, interviewers made no further attempt to locate potential respondents who could not be located by the central office. This process is analogous to the accepted practice of subsampling nonrespondents for more intensive effort—in this case, we subsampled cases in the clustered sample for field follow-up, but did not follow up unlocated cases in the unclustered sample. . When creating composite weights

---

[42] We obtained these totals from a frame file provided by SSA, giving information on basic demographics for all SSI and SSDI beneficiaries. The totals excluded 185,840 beneficiaries from Puerto Rico and outlying territories.

[43] These totals exclude 207 participants who resided in Puerto Rico or other outlying territories (the target population was limited to the 50 States and the District of Columbia). Of these 207 participants, 8 relied on the traditional payment system, 19 on SVRAs acting as ENs, and 180 on non–SVRA ENs.

[44] The central office is the Mathematica Survey Operations Center.

(described in the next section), we zeroed out the weights for the unlocated cases in the unclustered sample.[45] In Table III.5, we present the final sample sizes for the Ticket Participant Sample.

Table III.5. Survey Population and Initial Augmented and Final Sample Sizes, by Sampling Strata in the Participant Survey

Sampling Strata (Payment System/ Provider Type)	Survey Population[a]	Initial Augmented Sample Size[b]	Released Sample
Total Participants	85,038	11,863	4,334
1. Traditional payment type	68,592	3,069	1,083
2. Non–SVRA ENs		6,118	2,157
Clustered sample	12,728	2,818	1,049
Unclustered sample	12,728	3,300	1,108
In PSUs	3,084	788	273
Not in PSUs	9,644	2,512	835
3. SVRA ENs		2,676	1,094
Clustered sample	3,718	426	320
Unclustered sample	3,718	2,250	774
In PSUs	426	256	100
Not in PSUs	3,292	1,994	674

Source: Sample allocation and counts computed by Mathematica.

[a] This column reflects weighted totals before compositing. The totals exclude 207 participants who lived in Puerto Rico or other outlying territories (the target population was limited to the 50 states and the District of Columbia).

[b] The initial (augmented) and final (released) sample sizes include participants for whom the number obtained from the original sample design was insufficient for analysis. For participants using either SVRAs acting as ENs or non-SVRA ENs, we used a paired sample design, whereby participants in the PSUs could potentially be selected for both samples.

As indicated, for the clustered samples for Ticket participants (Traditional, Non-SVRA EN clustered, and SVRA EN clustered), we allocated the sample across the 79 PSUs, with the Los Angeles PSU receiving a double allocation because it had two selections. Given the smaller population sizes for Ticket participants when compared to the broader beneficiary population, we used only the full PSUs; we did not use the 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 Weights

We computed the initial weights 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, among participants with Tickets assigned to SVRAs functioning as ENs or to non–SVRA ENs, we computed the initial weights based on the selection probability within the two sampling strata (in one PSU or not in any PSU). With only a portion of the augmented sample released for use, we then adjusted the initial weights for the sample used in the survey.

[45] If a sample member selected as part of both the clustered and unclustered samples, was sent to the field for further follow-up and was then 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 "not selected for field follow-up." If such a case was duplicated in the clustered sample, the clustered sample case kept its original weight, appropriately adjusted so that the sum of the weights remained the same.

2. Dual-Frame Estimation

To obtain estimates for the SVRA and non–SVRA Ticket Participant subsamples, we had to use a "paired sample design" that combined the clustered and unclustered samples while accounting for different follow-up rules. The design required the creation of composite weights for application to the combined samples. As noted, if the central office could not locate a sample member in the unclustered sample, the office determined that the individual was "not selected for field followup" and thus undertook no further locating efforts. However, if the central office could not locate a sample member in the clustered sample, the case went to the field for additional locating efforts (field follow-up).

a. Conceptual Framework for Composite Weights

Consider a survey estimate, *Est(Y)*, such as the proportion currently working, that is computed using information from two independent samples, such as the clustered and unclustered samples described above. To compute this estimate, the two samples may not be combined without first adjusting the weights because the clustered and unclustered samples in the Ticket Participant Sample represent the same target population among Ticket participants. Separate estimates may be computed from each sample, within each payment type, and then combined by using the following equation:

(1) $$\mathrm{Est}(\mathrm{Y}) = \lambda \mathrm{Y}(\mathrm{clustered}) + (1-\lambda)\,\mathrm{Y}(\mathrm{unclustered})$$

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 participants with Tickets assigned to SVRAs functioning as ENs in the Round 4 data, the clustered sample accounted for 232 respondents and the unclustered sample for 446 respondents. The estimates to be combined are the proportion of the 232 in the clustered sample who are currently working and the proportion of the 446 in the unclustered sample who are currently working. In practice, of course, the calculation is more complicated because we need to account for the different rules used in the two samples for following up with nonrespondents or unlocated sample members (discussed later). For the sampling variance, *V(Y)*, the estimate is computed with the following equation:

(2) $$\mathrm{V}(\mathrm{Y}) = \lambda 2\mathrm{V}(\mathrm{Y}(\mathrm{clustered})) + (1-\lambda)2\mathrm{V}(\mathrm{Y}(\mathrm{unclustered}))$$

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 may be calculated as:

(3) $$\begin{aligned}\lambda &= 1/\mathrm{V}\Big(\mathrm{Y}(\mathrm{clustered})/\big[1/\mathrm{V}(\mathrm{Y}(\mathrm{clustered})) + 1/\mathrm{V}(\mathrm{Y}(\mathrm{unclustered}))\big]\Big)\\ &= \mathrm{V}(\mathrm{Y}(\mathrm{unclustered}))/\big[\mathrm{V}(\mathrm{Y}(\mathrm{clustered})) + \mathrm{V}(\mathrm{Y}(\mathrm{unclustered}))\big]\end{aligned}$$

In this case, the minimum variance is:

$$\text{(4)} \quad \mathrm{V}\left(\mathrm{Y}\right)=\left[\mathrm{V}\left(\mathrm{Y}\left(\text{clustered}\right)\right)*\mathrm{V}\left(\mathrm{Y}\left(\text{unclustered}\right)\right)\right]/\left[\mathrm{V}\left(\mathrm{Y}\left(\text{clustered}\right)\right)+\mathrm{V}\left(\mathrm{Y}\left(\text{unclustered}\right)\right)\right]$$

To compute the combined-sample estimate with minimum variance, we derive survey estimates 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. While this process produces minimum variance estimates, it is computer-intensive and results in some inconsistencies among estimates for percentages and proportions because of different values of λ among levels of categorical variables.

Since Round 2, we have used an alternative approach that identifies a single lambda calculated by using sample sizes and design effects attributable to unequal weighting for the two samples. In particular, λ acts as a weighting factor, with more weight given to the larger sample. The formula for λ includes sample sizes adjusted for the design effect attributable to unequal weighting. The formula for λ follows:

$$\text{(5)} \qquad \lambda=\frac{n\left(clustered\right)/\,deff\left(clustered\right)}{n\left(clustered\right)/\,deff\left(clustered\right)+n\left(clustered\right)/\,deff\left(unclustered\right)}$$

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 attributable 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 may be separated into two parts: the portion requiring field follow-up and the portion not requiring field follow-up. For the latter portion (that is, those who may be located through the central office's locating efforts), both the clustered and unclustered samples are independent samples that can provide unbiased estimates for this subpopulation. However, for the portion of the target population requiring field follow-up (that is, those who may not be located through the central office's locating efforts), only the clustered sample can provide unbiased estimates for this subpopulation because unclustered sample cases were not eligible for field follow-up.

For the subpopulation that may be located by the central office, the clustered and unclustered samples may be combined by using the compositing method (called a "dual frame" estimation procedure). The following equation computes the composite weight for each sample member in the clustered central office–located sample:

$$\text{(6)} \qquad WT=\lambda\ WT\left(\text{clustered central office-located sample weight}\right)$$

For units in the unclustered central office–located sample, the following equation computes the composite weight for each sample member in the unclustered central office–located sample:

$$\text{(7)} \qquad WT=\left(1-\lambda\right)WT\left(\text{unclustered central office-located sample weight}\right)$$

Conversely, for the subpopulation of persons not found through the central office's locating efforts, only the clustered sample may be used. In this case, no combining is required, and we used the clustered weight directly as follows:

$$WT = 1 * WT\left(\text{clustered field-located sample weight}\right) \quad (8)$$

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

With the paucity of sample members in the PSUs in some cases, the unclustered sample was often much larger than the clustered sample. Therefore, combining samples and creating composite weights sometimes resulted in weights with unacceptably high levels of variation and necessitated trimming to reduce such variation (described later).

3. Nonresponse Adjustment

As with the Representative Beneficiary Survey, we adjusted the sampling weights in two stages: one stage for sample members who could not be located and another stage for those who, once located, refused to respond. For the Ticket Participant Sample, we calculated the nonresponse adjustments (including both the location and cooperation adjustments) for all three provider-payment-type subpopulations by using logistic propensity models. For participants with Tickets assigned to either SVRAs functioning as ENs or non–SVRA ENs, we applied the nonresponse adjustments to the composite weights for the clustered and unclustered samples. Roughly equal sample sizes with vastly different population sizes for the three provider-payment types resulted in substantial differences in the magnitude of the weights. Thus, we calculated separate adjustments for each of the three subpopulations, first for the location adjustment and subsequently for the cooperation adjustment. The result was six weight adjustments, including the three location adjustments for the three participant subpopulations, and three cooperation adjustments for the same three subpopulations, by using logistic propensity models. The models were fitted in the same way as the adjustment models for the Representative Beneficiary Sample (Section B.2 of this chapter).

As with the Representative Beneficiary Sample, we wanted to limit the value of the location adjustment to less than 2.0 and the value of the response adjustment to 3.0. We defined a single trimming class for each model.[46] The main factors or attributes affecting our ability to locate and interview Ticket Participant sample members included the same factors as those used to locate and interview Representative Beneficiary sample members: personal characteristics of the sample member (race, ethnicity, gender, and age), identity of the payee with respect to the beneficiary, whether the beneficiary and the applicant for benefits lived in the same location, how many phones or addresses are in the SSA files for the beneficiary, beneficiary's living situation, and geographic characteristics, including attributes of the county where the beneficiary resides. In addition, the following factors or attributes affected our ability to locate and interview Ticket Participant Sample

[46] Across the three Ticket participant subpopulations, we trimmed 11 location adjustment factors and 4 cooperation adjustment factors (details in Section C.2.d of this chapter).

members: type of beneficiary (recipient of SSI, SSDI, or both), primary disability, and type of disability claim (a person with a disability, a survivor, or other). In subsequent sections, we describe how the specific covariates for each of the six weight adjustments varied.

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.[47]

b. Response Rates

The 71.4 percent response rate for the Ticket Participant Sample is the product of the weighted location rate and weighted completion rate among located sample members. The weighted location rate is 93.1 percent, and the weighted cooperation rate (the weighted completion rate among located sample members) is 76.6 percent. Analogous to the Representative 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.

c. Factors Related to Location and Response

In Tables III.6 through III.8, we provide information on selected factors associated with locating a sample member within each of the three provider-payment-type subpopulations and the factors associated with the response among located sample members. The tables include unweighted counts of all sample members, counts of located sample members, and counts of sample members from whom we obtained a completed interview or whom we deemed ineligible. The tables also include the weighted location rate, weighted completion rate among located sample members, and weighted overall completion rate for these factors, which helped inform the decision about the final set of variables to be used to define the weighting classes and to be applied in the nonresponse adjustment models.

---

[47] Disposition codes 420 (institutionalized) and 430 (unavailable during field period) were classified as nonrespondent codes in Round 4, even though they were considered ineligible codes in Round 1. This classification affected one case in the Round 4 Ticket Participant Sample. As a result, the nonresponse adjusted weight for the case was 0 in Round 4, 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 was noticeably small.

Table III.6. Weighted Location and Response Rates for the Ticket Participant Sample, SVRA ENs, by Selected Characteristics

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
All	958[a]	909	90.0	693	77.0	69.0
SSI Only, SSDI Only, or Both SSI and SSDI						
SSI only	247	228	87.0	173	77.6	67.5
SSDI only	437	417	90.6	317	75.5	68.1
Both SSI and SSDI	274	264	91.9	203	78.7	71.7
SSI or SSDI						
SSI only or both SSI and SSDI	521	492	89.6	376	78.2	69.7
SSDI only or both SSI and SSDI	711	681	91.1	520	76.8	69.5
Constructed Disability Status						
Deaf	31	28	91.3	13	43.2	40.2
Mental	610	578	88.4	453	78.9	69.3
Physical	310	296	92.9	222	76.8	70.9
Unknown	7	7	100.0	5	71.1	65.4
Beneficiary's Age (four categories)						
18 to 29	271	255	91.9	188	76.3	69.4
30 to 39	177	164	81.4	124	72.5	59.4
40 to 49	237	228	94.1	175	76.5	72.1
50 and older	273	262	90.9	206	81.0	72.7
Sex						
Male	496	473	91.0	366	78.8	71.2
Female	462	436	89.0	327	75.0	66.6
Hispanicity						
Hispanic	7	7	100.0	5	87.2	88.2
Non–Hispanic/unknown	951	902	90.0	688	76.9	68.9
Race						
White	626	598	91.8	452	75.8	69.6
Black	211	196	83.1	159	80.7	65.9
Unknown	110	105	92.5	73	74.3	68.0
Asian American, Pacific Islander,	3	3	100.0	2	74.0	71.4
North American Indian, or Alaskan Native	8	7	92.6	7	100.0	92.6
Living Situation						
Living alone	608	572	88.6	432	77.2	68.0
Living with others	69	67	92.8	50	75.2	68.7
In institution or unknown	281	270	92.6	211	77.0	71.2
Did the Applicant for Benefits Live in Same ZIP Code as the Beneficiary?						
No	79	71	77.2	56	80.9	61.9
Yes	646	616	90.6	471	78.2	70.3
No information	233	222	94.5	166	71.8	68.1

Table III.6 *(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	49	44	82.4	36	81.3	65.9
Payee is a family member	265	249	91.9	196	80.5	73.6
Payee is an institution	78	76	97.4	58	77.6	75.2
Other	566	540	88.7	403	75.0	66.2
Count of Phone Numbers in File						
Only one phone number in file	18	18	100.0	15	86.5	86.1
Two phone numbers in file	154	148	93.9	110	76.7	70.6
Three phone numbers in file	62	55	79.1	38	69.6	55.0
Four phone numbers in file	62	55	87.0	39	77.0	65.6
Five or more phone numbers in file	140	125	76.2	88	72.5	54.5
No information	522	508	97.0	403	79.7	77.2
Count of Addresses in File						
Only one address in file	544	527	94.0	415	78.9	73.8
Two addresses in file	247	231	87.2	172	75.4	65.0
Three or more addresses in file	70	56	71.6	33	68.2	48.1
No information	97	95	96.0	73	79.4	76.1
Type of Claim						
Survivor	65	62	90.7	46	72.4	65.6
Disabled	653	626	91.3	480	77.4	70.3
Unknown	240	221	86.3	167	77.1	66.3
Census Region						
Midwest	696	659	91.6	505	76.9	70.4
Northeast	115	108	93.5	78	72.2	67.8
South	140	135	81.1	105	81.0	64.2
West	7	7	100.0	5	69.1	67.5
Census Division						
East North Central	385	369	93.6	287	77.7	72.8
East South Central	5	5	100.0	4	78.8	80.0
Middle Atlantic	2	2	100.0	2	100.0	100.0
Mountain	5	5	100.0	3	56.8	58.7
New England	113	106	93.4	76	71.7	67.1
Pacific	2	2	100.0	2	100.0	100.0
South Atlantic	103	98	76.9	76	81.1	60.6
West North Central	311	290	89.9	218	76.1	68.3
West South Central	32	32	100.0	25	80.9	80.4
Metropolitan						
Metropolitan areas of 1 million population or more	303	290	85.6	220	75.8	63.4
Metropolitan areas of 250,000 to 999,999 population	268	250	85.8	191	75.3	64.8
Metropolitan areas of fewer than 250,000 population	133	127	94.8	96	79.2	74.7
Nonmetropolitan areas adjacent to large metropolitan areas	24	22	91.5	17	75.7	70.8
Nonmetropolitan areas adjacent to medium or small metropolitan areas	124	120	96.9	91	77.3	74.8

Table III.6 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Nonmetropolitan areas not adjacent to metropolitan areas	106	100	94.1	78	80.4	75.5
County with Low Education						
Yes	34	31	55.1	27	93.0	50.0
No	924	878	92.6	666	76.0	70.3
County with Housing Stress						
Yes	86	81	73.0	63	82.5	57.9
No	872	828	92.4	630	76.3	70.5
Population Loss County						
Yes	210	202	94.2	157	75.7	71.2
No	748	707	89.3	536	77.2	68.6
Retirement Destination County						
Yes	54	52	96.5	41	77.5	75.6
No	904	857	89.6	652	77.0	68.5
Service- Dependent Economy County						
Yes	269	260	95.3	193	72.1	68.3
No	689	649	88.5	500	78.4	69.2
Nonspecialized- Dependent Economy County						
Yes	325	300	87.4	225	75.4	65.8
No	633	609	91.4	468	77.9	70.7
Government- Dependent Economy County						
Yes	68	64	94.1	56	86.6	82.4
No	890	845	89.7	637	76.2	67.9
County Racial/Ethnic Profile						
County with at least 90% non-Hispanic white population	363	333	87.9	254	76.9	67.6
County with plurality or majority Hispanic population	5	5	100.0	3	54.8	54.4
County with majority but fewer than 90% non-Hispanic white population	369	359	97.1	270	76.0	74.0
County with a racially/ethnically mixed population, no majority group	200	191	79.9	151	80.1	62.4
County with plurality or majority non-Hispanic black population	19	19	100.0	14	74.3	73.7
County with at least 20% American Indian population	2	2	100.0	1	49.9	50.0
Phase						
Phase 1	220	208	93.8	157	77.6	72.4
Phase 2	203	195	85.3	148	78.7	66.1
Phase 3	535	506	90.4	388	75.8	68.5

Source: NBS, Round 4.

[a] Total does not include 136 unclustered cases that were not followed up in the field.

Table III.7. Weighted Location and Response Rates for the Ticket Participant Sample, Non- SVRA ENs, by Selected Characteristics

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
All	1,939 [a]	1,824	93.2	1,399	76.1	71.3
SSI Only, SSDI Only, or Both SSI and SSDI						
SSI only	425	385	89.2	283	74.4	66.7
SSDI only	1,048	1,002	94.9	794	78.9	75.3
Both SSI and SSDI	466	437	92.9	322	71.4	66.5
SSI or SSDI						
SSI only or both SSI and SSDI	891	822	91.2	605	72.8	66.6
SSDI only or both SSI and SSDI	1,514	1,439	94.3	1,116	76.6	72.6
Constructed Disability Status						
Deaf	18	14	79.3	7	48.7	38.8
Mental	962	894	91.7	664	73.8	68.2
Physical	940	897	94.7	712	78.7	74.8
Unknown	19	19	100.0	16	84.0	83.8
Beneficiary's Age (four categories)						
18 to 29	332	305	92.2	220	72.9	67.5
30 to 39	370	345	91.5	258	73.6	68.0
40 to 49	503	479	93.8	365	74.7	70.3
50 and older	734	695	94.1	556	80.0	75.5
Sex						
Male	992	927	92.4	709	76.5	71.0
Female	947	897	93.9	690	75.8	71.6
Hispanicity						
Hispanic	98	93	92.2	75	80.3	75.3
Non-Hispanic/unknown	1,841	1,731	93.2	1,324	75.9	71.1
Race						
White	895	848	94.5	654	76.2	72.7
Black	700	654	91.9	500	77.1	70.8
Unknown	326	306	92.1	235	74.2	68.9
Asian American, Pacific Islander,	16	14	91.2	8	60.1	53.8
North American Indian, or Alaskan Native	2	2	100.0	2	100.0	100.0
Living Situation						
Living alone	1,091	1,017	92.6	770	75.3	70.2
Living with others	131	121	89.4	87	71.4	63.9
Living with parents	6	5	79.1	2	50.9	41.7
In institution or unknown	711	681	94.8	540	78.5	74.7
Did the Applicant for Benefits Live in Same ZIP Code as the Beneficiary?						
No	201	184	90.3	129	71.9	65.0
Yes	1,252	1,175	92.8	903	76.2	71.0
No information	486	465	95.4	367	77.8	74.7

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	93	84	89.0	68	81.6	72.9
Payee is a family member	355	335	93.4	239	70.5	66.3
Payee is an institution	57	54	93.4	38	69.6	65.8
Other	1,434	1,351	93.3	1,054	77.5	72.7
Count of Telephone Numbers in File						
Only one phone number in file	19	19	100.0	14	78.0	77.7
Two phone numbers in file	280	271	97.3	209	75.6	73.6
Three phone numbers in file	145	133	90.1	92	68.1	61.6
Four phone numbers in file	101	91	91.6	64	68.7	63.7
Five or more phone numbers in file	399	340	84.4	250	73.3	62.3
No information	995	970	97.6	770	80.3	78.4
Count of Addresses in File						
Only one address in file	1,012	975	96.2	780	79.9	76.9
Two addresses in file	639	602	93.2	442	72.4	67.7
Three or more addresses in file	172	134	76.6	83	65.8	50.6
No information	116	113	97.7	94	85.4	83.3
Type of Claim						
Survivor	76	70	91.3	51	74.2	68.6
Disabled	1,457	1,387	94.4	1,078	76.8	72.8
Unknown	406	367	89.2	270	74.2	66.4
Census Region						
Midwest	321	306	94.5	240	79.6	75.3
Northeast	326	307	92.8	234	75.2	70.7
South	805	755	92.8	586	76.7	71.4
West	487	456	93.1	339	72.9	68.4
Census Division						
East North Central	255	243	94.2	191	79.0	74.5
East South Central	91	88	94.2	72	78.6	74.3
Middle Atlantic	199	192	95.6	150	80.3	76.6
Mountain	143	138	95.5	104	72.4	69.4
New England	127	115	88.3	84	67.1	61.1
Pacific	344	318	91.9	235	73.1	67.9
South Atlantic	535	500	92.5	380	75.1	69.7
West North Central	66	63	95.2	49	81.4	77.6
West South Central	179	167	92.8	134	80.1	74.5
Metropolitan						
Metropolitan areas of 1 million population or more	1,237	1,163	93.0	879	75.9	70.8
Metropolitan areas of 250,000 to 999,999 population	472	438	91.3	340	73.9	68.5
Metropolitan areas of fewer than 250,000 population	118	113	94.4	89	78.5	74.1
Nonmetropolitan areas adjacent to large metropolitan areas	26	25	94.1	22	83.7	78.2
Nonmetropolitan areas adjacent to medium or small metropolitan areas	53	53	100.0	42	78.8	79.2

Table III.7 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Nonmetropolitan areas not adjacent to metropolitan areas	33	32	97.9	27	83.6	81.7
County with Low Education						
Yes	337	312	92.2	233	74.8	69.2
No	1,602	1,512	93.4	1,166	76.4	71.7
County with Housing Stress						
Yes	1,099	1,029	92.9	776	75.2	70.1
No	840	795	93.5	623	77.1	72.6
Population Loss County						
Yes	212	201	93.9	149	77.4	72.2
No	1,727	1,623	93.1	1,250	76.0	71.2
Retirement Destination County						
Yes	240	231	94.6	180	76.3	72.6
No	1,699	1,593	92.9	1,219	76.1	71.1
Service- Dependent Economy County						
Yes	1,195	1,113	91.7	838	73.6	67.7
No	744	711	94.9	561	79.2	75.6
Nonspecialized- Dependent Economy County	321	305	94.8	250	83.5	79.3
Yes	1,618	1,519	92.8	1,149	74.4	69.4
No						
Government- Dependent Economy County						
Yes	184	175	95.3	138	78.9	75.5
No	1,755	1,649	92.9	1,261	75.8	70.8
County Racial/Ethnic Profile						
County with at least 90% non-Hispanic white population	106	103	96.9	82	81.4	79.2
County with plurality or majority Hispanic population	278	261	93.1	198	74.8	70.3
County with majority but fewer than 90% non-Hispanic white population	621	589	93.3	464	76.6	71.9
County with a racially/ethnically mixed population, no majority group	813	761	93.0	572	74.8	69.8
County with plurality or majority non-Hispanic black population	119	108	88.5	81	78.0	68.6
County with at least 20% American Indian population	2	2	100.0	2	100.0	100.0
Phase						
Phase 1	713	669	93.0	510	75.3	70.2
Phase 2	408	387	93.2	310	79.3	74.6
Phase 3	818	768	93.3	579	75.0	70.2

Source: NBS, Round 4.

[a] Total does not include 218 unclustered cases that were not followed up in the field.

Table III.8. Weighted Location and Response Rates for Ticket Participant Sample, Traditional Payment System, by Selected Characteristics

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
All	1,083	1,009	93.3	765	76.3	71.5
SSI Only, SSDI Only, or Both SSI and SSDI						
SSI only	373	343	92.1	257	75.8	70.0
SSDI only	434	406	93.5	313	77.5	72.9
Both SSI and SSDI	276	260	94.3	195	74.9	71.1
SSI or SSDI						
SSI only or both SSI and SSDI	649	603	93.1	452	75.4	70.5
SSDI only or both SSI and SSDI	710	666	93.8	508	76.5	72.2
Constructed Disability Status						
Deaf	47	41	88.2	25	61.2	55.2
Mental	623	584	93.8	443	76.0	71.8
Physical	400	372	93.0	285	77.5	72.1
Unknown	13	12	93.0	12	100.0	93.0
Beneficiary's Age (four categories)						
18 to 29	363	335	92.3	258	77.3	72.0
30 to 39	171	156	91.4	111	71.4	65.4
40 to 49	273	257	94.2	201	78.7	74.4
50 and older	276	261	94.7	195	75.5	71.6
Sex						
Male	596	557	93.5	424	76.9	72.1
Female	487	452	93.0	341	75.5	70.8
Hispanicity						
Hispanic	75	67	89.7	50	75.8	67.7
Non-Hispanic/unknown	1,008	942	93.5	715	76.3	71.7
Race						
White	630	586	93.1	435	74.9	69.8
Black	199	189	95.0	157	84.3	80.1
Unknown	244	226	92.9	168	74.1	69.8
Asian American, Pacific Islander,	6	5	84.7	4	73.4	62.3
North American Indian, or Alaskan Native	4	3	74.3	1	27.8	24.8
Living Situation						
Living alone	686	630	92.0	476	76.0	70.2
Living with others	116	113	97.3	85	75.6	74.1
Living with parents	11	10	90.6	7	73.4	66.0
In institution or unknown	270	256	94.8	197	77.3	73.8
Did the Applicant for Benefits Live in Same ZIP Code as the Beneficiary?						
No	123	113	92.3	76	67.6	62.6
Yes	713	668	93.8	517	77.8	73.4
No information	247	228	92.2	172	76.0	70.3

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	45	42	93.0	33	77.8	73.1
Payee is a family member	364	348	95.7	261	76.0	73.0
Payee is an institution	67	58	86.8	45	78.8	67.3
Other	607	561	92.6	426	76.0	70.9
Count of Phone Numbers in File						
Only one phone number in file	21	21	100.0	17	81.4	81.7
Two phone numbers in file	168	160	95.5	130	81.5	78.1
Three phone numbers in file	90	85	94.5	65	77.6	74.0
Four phone numbers in file	84	71	84.1	54	76.5	64.5
Five or more phone numbers in file	227	192	84.6	145	75.3	64.5
No information	493	480	97.4	354	74.4	72.7
Count of Addresses in File						
Only one address in file	588	571	97.2	464	82.2	79.9
Two addresses in file	325	298	91.7	215	73.2	66.7
Three or more addresses in file	125	98	79.0	58	59.6	47.3
No information	45	42	93.0	28	67.2	62.4
Type of Claim						
Survivor	78	75	96.3	52	71.7	68.4
Disabled	642	601	93.6	465	77.4	72.9
Unknown	363	333	91.9	248	75.3	69.5
Census Region						
Midwest	234	216	92.5	171	78.9	73.6
Northeast	196	183	93.7	134	74.6	70.1
South	381	360	94.5	269	75.7	71.6
West	272	250	92.0	191	75.8	70.2
Census Division						
East North Central	168	157	93.6	124	78.3	74.2
East South Central	55	54	98.0	33	63.8	63.1
Middle Atlantic	129	117	90.8	79	69.1	63.0
Mountain	75	69	92.0	54	76.7	71.9
New England	67	66	98.5	55	83.1	82.0
Pacific	197	181	91.9	137	75.5	69.5
South Atlantic	200	184	92.2	140	76.7	70.7
West North Central	66	59	89.7	47	80.4	72.1
West South Central	126	122	96.7	96	79.3	76.8
Metropolitan						
Metropolitan areas of 1 million population or more	451	422	93.6	316	75.3	70.7
Metropolitan areas of 250,000 to 999,999 population	293	267	91.4	202	76.1	69.7
Metropolitan areas of fewer than 250,000 population	204	193	94.5	148	77.8	73.9
Nonmetropolitan areas adjacent to large metropolitan areas	54	51	94.7	42	82.5	78.4
Nonmetropolitan areas adjacent to medium or small metropolitan areas	29	28	96.8	23	79.0	79.2

Table III.8 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Nonmetropolitan areas not adjacent to metropolitan areas	52	48	92.2	34	71.5	66.8
County with Low Education						
Yes	146	137	93.8	93	68.4	64.2
No	937	872	93.2	672	77.4	72.5
County with Housing Stress						
Yes	515	479	93.0	363	76.2	71.2
No	568	530	93.5	402	76.3	71.7
Population Loss County						
Yes	97	87	90.3	61	69.9	63.7
No	986	922	93.5	704	76.9	72.2
Retirement Destination County						
Yes	160	147	91.7	108	74.1	68.1
No	923	862	93.5	657	76.6	72.0
Service- Dependent Economy County						
Yes	500	464	92.9	352	76.4	71.4
No	583	545	93.5	413	76.2	71.6
Nonspecialized- Dependent Economy County						
Yes	289	265	91.6	195	74.0	68.0
No	794	744	93.8	570	77.1	72.7
Government- Dependent Economy County						
Yes	121	113	93.4	88	79.3	74.5
No	962	896	93.2	677	75.9	71.1
County Racial/Ethnic Profile						
County with at least 90% non-Hispanic white population	138	128	93.0	103	79.7	75.0
County with plurality or majority Hispanic population	135	125	92.2	91	72.2	67.2
County with majority but fewer than 90% non-Hispanic white population	498	461	92.7	350	76.3	71.0
County with a racially/ethnically mixed population, no majority group	297	282	95.1	212	76.7	72.9
County with plurality or majority non-Hispanic black population	12	11	91.5	7	59.4	58.4
County with at least 20% American Indian population	3	2	62.0	2	100.0	62.0
Phase						
Phase 1	344	318	92.6	235	74.8	69.4
Phase 2	306	283	92.5	211	75.3	70.1
Phase 3	433	408	94.3	319	78.1	73.9

Source: NBS, Round 4.

d. Propensity Models for Weight Adjustments

The weight adjustments used in the Ticket Participant Sample were based on predicted propensities from a logistic regression model. As indicated earlier, we calculated the adjustments by taking the inverse of the predicted location and cooperation propensities, which were determined by using separate logistic models for each of the three provider-payment-type subpopulations.

The adjusted weight for each sample case is the product of the initial sampling weight and the adjustment factor, trimmed to ensure that the maximum location adjustment did not exceed 2 and that the maximum cooperation adjustment did not exceed 3.

Below, we provide the primary factors used to calculate the location adjustments, with the potential levels used in the models. (Appendix D details how the levels were collapsed for each model.)

1. **DIVISION.** Geographic region of beneficiary's place of residence, based on U.S. Census divisions, with 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, (9) New England
2. **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 fewer 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 fewer than 1 million, (6) beneficiary lived in nonmetropolitan area not adjacent to any metropolitan area
3. **DIG.** Disability diagnostic classification; possible levels: (1) mental disability, (2) physical disability (excluding deaf cases), (3) deaf, (4) unknown
4. **LIVING.** Beneficiary's living situation; possible levels: (1) beneficiary lives alone, (2) beneficiary lives with his or her parents, (3) beneficiary lives in an institution, (4) information unknown
5. **AGECAT.** Beneficiary's age category; possible levels: (1) age 18 to 29, (2) age 30 to 39, (3) age 40 to 49, (4) age 50 to 64
6. **SSI_SSDI.** Beneficiary status; possible levels: (1) SSI only, (2) SSDI only, (3) both SSI and SSDI
7. **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, (3) an institution received payments on behalf of the beneficiary or identity of payee not known
8. **RACE.** Possible levels: (1) white, (2) black, (3) Asian or Pacific Islander, (4) not white, black, or Asian/Pacific Islander or unknown
9. **CNTYRACE.** County racial ethnic profile; two levels: (1) county with racially/ethnically mixed population based on 2000 Census, no majority group, (2) other racial/ethnic profile in county

10. **CNTYSVC.** Service-dependent economy county; two levels: (1) county with 45 percent or more of average annual labor and proprietors' earnings derived from services (SIC categories of retail trade; finance, insurance, and real estate; and services) during 1998–2000, (2) county without this attribute
11. **MOVE.** Count of addresses in SSA files; four levels: (0) no information, (1) one address in file, (2) two addresses in file, (3) three or more addresses in file
12. **PHONE.** Count of phone numbers in SSA files; three levels: (0) no information, (1) one phone number in file, (2) two or more phone numbers in file

In Table III.9, we list the variables used in each Ticket participant location model. Appendix D features an expanded form of Table III.9, that presents the specific levels of the main effects for each model, along with parameter estimates and their standard errors.

Table III.9. Variables Used in the Location Logistic Propensity Models: Ticket Participant Sample

Variables in Location Model for Participants Using SVRAs Acting as EN Provider
Main Effects
MOVE (COUNT OF ADDRESSES ON FILE)
PHONE (COUNT OF PHONE NUMBERS ON FILE)
LIVING (LIVING SITUATION)
METRO (METROPOLITAN STATUS OF COUNTY)
CNTYRACE (COUNTY RACIAL/ETHNIC PROFILE)
CNTYSVC (SERVICE-DEPENDENT ECONOMY COUNTY)
Variables in Location Model for Participants Using Non-SVRA ENs as EN Provider
Main Effects
DIVISION (CENSUS DIVISION)
DIG (DISABILITY)
LIVING (LIVING SITUATION)
MOVE (COUNT OF ADDRESSES IN FILE)
PHONE (COUNT OF PHONE NUMBERS IN FILE)
SSI_SSDI (RECIPIENT OF SSI, SSDI, OR BOTH)
CNTYRACE (COUNTY RACIAL/ETHNIC PROFILE)
CNTYSVC (SERVICE-DEPENDENT ECONOMY COUNTY)
Two-Factor Interaction
PHONE*CNTYRACE
DIG*MOVE
DIG*CNTYRACE
DIG*PHONE
Variables in Location Model for Participants Using Traditional Payment System
Main Effects
LIVING
SSI_SSDI
METRO
MOVE (COUNT OF ADDRESSES IN FILE)
PHONE (COUNT OF PHONE NUMBERS IN FILE)
REPREPAYEE (IDENTITY OF PAYEE WITH RESPECT TO BENEFICIARY)
AGECAT (AGE CATEGORY)
RACE
DIG (DISABILITY)

Below, we list the primary factors in the cooperation models, noting only the base variables with all possible levels. We provided some of the base variables in the discussion of location adjustments and do not repeat their earlier descriptions. (Appendix D describes how the levels were collapsed for each model.)

1. **MOVE.** Count of addresses in SSA files; four levels: (0) no information, (1) one address in file, (2) two addresses in file, (3) three or more addresses in file.
2. **DIG.** Disability diagnostic classification; possible levels: (1) mental disability, (2) physical disability (excluding deaf cases), (3) deaf, (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, (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; two levels: (1) beneficiary and applicant lived in the same ZIP code, (2) beneficiary and applicant lived in different ZIP codes/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 fewer 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 fewer than 1 million, (6) beneficiary lived in nonmetropolitan area not adjacent to any metropolitan area.
6. **GENDER (SEX).** Two levels: (1) male, (2) female.
7. **REGION or DIVISION.** Geographic region of beneficiary's place of residence: DIVISION is based on U.S. Census divisions, with 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, (9) New England; REGION is based on U.S. Census regions with four levels, which may be collapsed from the nine levels of DIVISION: (1) West is Pacific + Mountain, (2) Midwest is East North Central + West North Central, (3) South is East South Central + West South Central + South Atlantic, (4) Northeast is Middle Atlantic + New England.[48]
8. **LIVING.** Beneficiary's living situation; possible levels: (1) beneficiary lives alone, (2) beneficiary lives with his or her parents, (3) beneficiary lives in an institution, (4) information unknown.
9. **PHONE.** Count of phone numbers in SSA files; three levels: (0) no information, (1) one phone number in file, (2) two or more phone numbers in file.
10. **AGECAT.** Beneficiary's age category; possible levels: (1) age 18 to 29, (2) age 30 to 39, (3) age 40 to 49, (4) age 50 to 64.

[48] Many of the cooperation models used REGION instead of DIVISION. If a U.S. Census division was used in a model, then the U.S. Census region corresponding to that division could not be in the model.

11. **SSI_SSDI.** Beneficiary status; possible levels: (1) SSI only, (2) SSDI only, (3) both SSI and SSDI.
12. **TOC.** Type of claim; possible levels: (1) survivor claim, (2) disability claim, (3) type of claim unknown.
13. **RACE.** Possible levels: (1) white, (2) black, (3) Asian or Pacific Islander, (4) not white, black, or Asian/Pacific Islander or unknown.
14. **HISPANICITY.** Whether the beneficiary was Hispanic or not; two levels: (1) Hispanic, (2) not Hispanic or unknown.
15. **CNTYRACE.** County racial ethnic profile; two levels: (1) county with racially/ethnically mixed population based on 2000 Census, no majority group, (2) other racial/ethnic profile in count.
16. **CNTYPOPLOSS.** County with population loss; two levels: (1) county with population loss in both 1980–1990 and 1990–2000 decennial periods, (2) county with population gain in 1980–1990 and/or 1990–2000 decennial periods.
17. **CNTYLOWEDUC.** County with low education; two levels: (1) county where 25 percent or more of residents age 25 through 64 had neither a high school diploma nor Graduate Equivalency Degree (GED) in 2000, (2) county without this attribute.
18. **CNTYHOUSSTRESS.** County with issues related to housing; two levels: (1) 30 percent or more of households had one or more of these housing conditions in 2000: lacked complete plumbing, lacked complete kitchen, paid 30 percent or more of income for owner costs or rent, or had more than one person per room, (2) county without this attribute.
19. **CNTYGOV.** County with government-dependent economy: (1) 15 percent or more of average annual labor and proprietors' earnings derived from Federal and state government during 1998–2000, (2) county without this attribute.
20. **CNTYSVC.** Service-dependent economy county; two levels: (1) county with 45 percent or more of average annual labor and proprietors' earnings derived from services (SIC categories of retail trade; finance, insurance, and real estate; and services) during 1998–2000, (2) county without this attribute.
21. **CNTYNONSP.** Nonspecialized-dependent economy county; two levels: (1) county that did not meet economic thresholds for government-dependent economy, mining-dependent economy, manufacturing-dependent economy, farming-dependent economy, or service-dependent economy during 1998–2000, (2) county that meets one or more thresholds for the listed economic dependencies.

The models for the cooperation of sample members included various interactions among the above variables. In Table III.10, we list the variables included in each Ticket participant cooperation model. Appendix D features an expanded form of Table III.10, with levels appropriately collapsed for each model and the specific levels of the interactions, along with parameter estimates and their standard errors.

Table III.10. Variables in the Cooperation Logistic Propensity Models: Ticket Participant Sample

Variables in SVRA EN Cooperation Model
Main Effects
REPREPAYEE (IDENTITY OF PAYEE WITH RESPECT TO BENEFICIARY)
MOVE (COUNT OF ADDRESSES IN FILE)
PHONE (COUNT OF PHONE NUMBERS IN FILE)
SSI_SSDI (RECIPIENT OF SSI, SSDI, OR BOTH)
DIVISION (CENSUS DIVISION)
PHONE (COUNT OF PHONES IN FILE)
DIG (DISABILITY)
RACE
GENDER
AGECAT (AGE CATEGORY)
CNTYSVC (SERVICE-DEPENDENT ECONOMY COUNTY)
CNTYNONSP (NONSPECIALIZED-DEPENDENT ECONOMY COUNTY)
Two- Factor Interactions
AGECAT*CNTYNONSP
RACE*GENDER
DIG*AGECAT
GENDER*PHONE
Variables in Non- SVRA EN Cooperation Model
Main Effects
REPREPAYEE (IDENTITY OF PAYEE WITH RESPECT TO BENEFICIARY)
MOVE (COUNT OF ADDRESSES IN FILE)
PHONE (COUNT OF PHONE NUMBERS IN FILE)
GENDER
SSI_SSDI (RECIPIENT OF SSI, SSDI, OR BOTH)
REGION (CENSUS REGION)
PHONE (COUNT OF PHONES IN FILE)
DIG (DISABILITY)
RACE
HISPANICITY
PDZIPSAME (WHETHER APPLICANT AND BENEFICIARY LIVE IN SAME ZIP CODE)
CNTYRACE (COUNTY RACIAL/ETHNIC PROFILE)
CNTYSVC (SERVICE-DEPENDENT ECONOMY COUNTY)
CNTYGOV (GOVERNMENT-DEPENDENT ECONOMY COUNTY)
CNTYHOUSSTRES (COUNTY WITH ISSUES RELATED TO HOUSING)
CNTYNONSP (NONSPECIALIZED-DEPENDENT ECONOMY COUNTY)
METRO (METROPOLITAN STATUS OF COUNTY)
Two- Factor Interactions
PDZIPSAME*CNTYNONSP
PHONE*CNTYNONSP
FEMALE*PHONE
REGION*METRO
RACE*PHONE
CNTYRACE*CNTYSVC
RACE*CNTYHOUSSTRES
SSI_SSDI*PHONE
DIG*CNTYNONSP

Table III.10 *(continued)*

Variables in Traditional Cooperation Model
Main Effects
MOVE (COUNT OF ADDRESSES IN FILE)
PHONE (COUNT OF PHONE NUMBERS IN FILE)
DIVISION (CENSUS DIVISION)
DIG (DISABILITY)
TOC (TYPE OF DISABILITY CLAIM)
RACE
CNTYRACE (COUNTY RACIAL/ETHNIC PROFILE)
CNTYPOPLOSS (POPULATION LOSS COUNTY)
CNTYLOWEDUC (COUNTY WITH LOW EDUCATION)
Two-Factor Interactions
TOC*BLACK

As with the beneficiary sample, the model-fitting process proved to be complex. After identifying a smaller pool of main effects and interactions for potential inclusion in the final model, we used backward and forward stepwise logistic regressions in the SAS LOGISTIC procedure to evaluate statistically and identify a set of models from which to select the final model. Given that the SAS logistic regression procedure does not incorporate the sampling design, we used the logistic regression procedure in SUDAAN to make the final selection of covariates.

For selecting variables or interactions in the stepwise procedures, we again included variables or interactions with a statistical significance level (alpha level) of 0.30 or lower (instead of the usual 0.05). Once we identified the candidate list of main effects and interactions, we used a thorough model-fitting process to determine a parsimonious model with few very small propensities.

In Table III.9, we summarize the main effects used to calculate the location adjustments; in Table III. 10, we summarize the main effects and interactions in the models for cooperation among located sample members. In Table III.11, we provide the R-squared values for the six logistic models.

Table III.11. Unadjusted and Adjusted R-Squared Values for Logistic Propensity Models in Ticket Participant Cross-Sectional Samples

Model			
Payment System/ Provider-Payment Type	Location or Cooperation	Unadjusted R-Squared Value	Adjusted R-Squared Value
SVRA EN	Location	0.159	0.337
SVRA EN	Cooperation	0.084	0.128
Non-SVRA EN	Location	0.079	0.201
Non-SVRA EN	Cooperation	0.081	0.122
Traditional	Location	0.080	0.205
Traditional	Cooperation	0.076	0.114

The unadjusted R-squared value for the location models ranged from 0.079 to 0.159 (0.201 to 0.337 when rescaled to have a maximum of 1). The unadjusted R-squared value for the nonresponse models ranged from a low of 0.076 (0.114 when rescaled as above) to 0.084 (0.128 when rescaled). The values are similar to those observed for other response propensity modeling efforts that used logistic regression with design-based sampling weights. In Table III.12, we present the percentages of concordant and discordant pairs and the p-values for the Hosmer-Lemeshow goodness-of-fit test.

Table III.12. Percentages of Concordant and Discordant Pairs and Hosmer-Lemeshow p-Values for Logistic Propensity Models in Ticket Participant Cross-Sectional Samples

Model				
Payment System/ Provider-Payment Type	Location or Cooperation	Percentage Concordant	Percentage Discordant	Hosmer-Lemeshow p-Value
SVRA EN	Location	74.3	23.3	0.244
SVRA EN	Cooperation	65.5	34.0	0.848
Non-SVRA EN	Location	78.0	21.0	0.462
Non-SVRA EN	Cooperation	65.0	34.5	0.548
Traditional	Location	79.9	19.2	0.902
Traditional	Cooperation	67.1	30.9	0.364

The minimum difference between the percentages of concordant and discordant pairs is 30.5 percentage points (the non–SVRA cooperation model). In general, the proportions of concordant and discordant pairs indicate stronger models for the location models compared to the cooperation models. The minimum p-value associated with the Hosmer-Lemeshow goodness-of-fit test is 0.244, indicating no evidence of lack of fit for any of the models.

4. Trimming

As indicated earlier, we trimmed adjustment factors so that the location adjustment factors did not exceed 2 and the cooperation adjustment factors did not exceed 3. In Table III.13, we provide the adjustment factors for all six logistic regression models before and after trimming as well as the number of adjustment factors trimmed.

Table III.13. Count of Trimmed Adjustment Factors and Range of Adjustment Factors Before and After Trimming

Model				
Payment System/ Provider-Payment Type	Location or Cooperation	Count of Number Trimmed	Range Before Trimming	Range After Trimming
SVRA EN	Location	1	1.00–2.02	1.00–2.00
SVRA EN	Cooperation	0	1.03–2.59	1.03–2.59
Non–SVRA EN	Location	10	1.00–2.49	1.00–2.00
Non–SVRA EN	Cooperation	2	1.04–3.50	1.04–3.00
Traditional	Location	0	1.00–1.97	1.00–1.97
Traditional	Cooperation	2	1.01–3.43	1.01–3.00

After we applied the trimmed adjustments to the sampling weights, we reviewed the distribution of weights to determine the need for trimming such weights. In view of the wide variation in the magnitude of the weights, which was attributable to the use of composite weights in the SVRA and non–SVRA provider-payment types, trimming was sometimes warranted in order to increase the survey estimates' precision. However, to reduce the potential for bias in the estimates, we minimized the extent of trimming. In Table III.14, we present the design effects attributable to unequal weighting associated with each of the six-phase/payment-type combinations before and after trimming, before post-stratification. We calculated design effects separately within trimming strata, which, in turn, we defined within the three strata based on payment system and provider type. In general, we defined the trimming strata according to whether the observation was in the clustered or unclustered sample. For unclustered cases, we further subdivided the trimming strata according to whether the sample case was/was not in a PSU. Table III.14 indicates the strata within which trimming was employed. In the absence of trimming for a payment system andprovider type, the table describes the maximum design effect across all trimming strata. In such an instance, the table does not present the stratum associated with that maximum design effect; in most cases, when no trimming is required, the design effects do not differ significantly across trimming strata.

Table III.14. Design Effects Attributable to Unequal Weights Before and After Trimming, Within Trimming Strata, for Payment Types in the Round 4 Ticket Participant Samples

		Design Effect Attributable to Unequal Weights	
Payment Type	Trimming Stratum in Which Trimming Occurred	Before Trimming	After Trimming
SVRA EN	No trimming (three trimming strata)	2.62 (maximum)	2.62 (maximum)
Non-SVRA EN	Clustered	1.63	1.60
Traditional	No trimming (three trimming strata)	1.08 (maximum)	1.08 (maximum)

Design effect attributable to unequal weights = $n\Sigma w^2 / \left(\Sigma w\right)^2$

5. Post-Stratification

After the nonresponse adjustment and trimming, we post-stratified the weights to the population age and gender totals for each payment type obtained from the SSA sampling frame. The sampling frame included all SSI or SSDI beneficiaries for each provider-payment type within the population of Ticket Participants. We rechecked the distributions of weights within each provider-payment type to determine the need for more weight trimming. We found no extreme weights after post-stratification.

IV. IMPUTATIONS

The NBS data collection instruments were administered with computer-assisted interviewing (CAI) technology. The technology allows the use of automated routing to move the respondent to the applicable questions and performs checks of the entered data for consistency and reasonableness. In addition, it does not permit a question to be left blank; therefore, the interviewer may not proceed until 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, although some item nonresponse will persist as when a question was mistakenly not asked and when "don't know" or "refused" were recorded as responses.

For the NBS, we used primarily two methods of imputation to compensate for item nonresponse: deductive (or logical) imputation and unweighted hot-deck imputation. However, for some variables, the data were insufficient for use of either method and thus required the use of specialized imputation procedures were employed to use with the available data. Selection of the methods was 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.

Deductive, or logical, imputation is based on a review of the data related to the imputed variable. It assigns a value that may be deduced from other data or for which there is a high degree of certainty that the value is correct.

The hot-deck imputation procedure involves 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. Ideally, the imputation class should contain sufficient sample members to avoid the selection of a single donor for several sample members with missing data.

The hot-deck procedure is computationally efficient, and, in a 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. The USDE study evaluated imputation methods in terms of bias of the mean, median, quartile, and variance estimates, coverage probability, confidence interval width, and average imputation error.

Although the variance of estimates was a key item used to evaluate methods by the USDE study, we made no attempt in this study to estimate the component of variance attributable to imputation, even though such a component is always positive. Users should be aware that variance estimates that use 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; thus, the component of variance would be very small in most cases.

For the NBS, the hot-deck imputation procedure used an unweighted selection process to select a donor, with selections made within imputation classes defined by key related variables for each application. In addition to the variables defining the imputation classes, we included a sorting variable that sorted the recipient and all donors within the imputation class together by levels of the variable. Using the sorted data within the imputation class, we randomly selected as the donor with equal probability a case immediately preceding or following a sample member with missing data. Therefore, the hot-deck procedure was unweighted and sequential, with a random component. 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. Given that the extent of missing values was very low for most variables, we used only a few donors more than once.[49]

Where appropriate, we made imputed values consistent with pre-existing nonmissing variables by excluding donors with potentially inconsistent imputed values. After processing each imputation, we used a variety of quality control procedures to evaluate the imputed values. If the initial imputed value was beyond an acceptable range or inconsistent with other data for that case, we repeated the imputation until the imputed value was in range and consistent with other reported data.

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 NBS. In addition, the imputation classes needed to possess a sufficient count of donors for each sample member with missing data. We used a variety of methods to form the imputation classes: bivariate cross-tabulations, step-wise regressions, and multivariate procedures such as CHAID.[50] To develop the imputation classes, we used information from both the interview and SSA data files. The classing and sorting variables were closely related to the variable to be imputed (the response variable). The sorting variables were either less closely related to the response variable than were the classing variables or were forms of the classing variables with finer levels. As an example of the latter situation, we sometimes used four age categories as imputation classes: (1) 18 to 29, (2) 30 to 39, (3) 40 to 49, and (4) 50 to 64. We could then use the actual age as a sorting variable to ensure that donors and recipients were as close together in age as possible.

In the case of missing values in the variables used to define imputation classes, we applied two strategies: (1) matching recipients to donors also missing the value for the covariate or (2) employing separate hot decks, depending on the availability of the variables defining the imputation classes. In the first instance, we treated the level defined as the missing value as a separate level. In other words, if a recipient was missing a value for a variable defining an imputation class, the donor also was missing the value for that variable. We used the first strategy if a large number of donors and recipients were missing the covariate in question. In the second instance, we used a variable for a given recipient to define the imputation class for that recipient only if there was no missing value for that variable. The variables used to define an imputation class for each recipient depended on what values were nonmissing among those variables.

The hot-deck software automatically identified situations in which the imputation class contained only recipients and no donors. In such cases, we collapsed imputation classes and once

[49] Household income, used to determine the federal poverty threshold indicator, was the exception. Approximately 15 percent gave no household income information at all, and an additional (approximately) 17 percent gave only general categories of income. Detailed levels of missingness are given for all imputed variables later in this chapter.

[50] Chi-Squared Automatic Interaction Detection software is attributed to Kass [1980] and Biggs et al. [1991], and its application in SPSS is described in Magidson [1993].

again performed the imputation with 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. If several covariates aided in imputing a given variable, the covariates less closely related to the variable requiring imputation were more likely than the important covariates in the imputation to have levels that we had to collapse. In addition, variables with a large number of levels also were more likely to have levels that we had to collapse. In general, if more than a very small number of imputation classes required collapsing, we dropped one or more variables from the definition of the imputation class and re-ran the imputation procedure.

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 impose the recoding of the constructed variable on these imputed values, rather than imputing the constructed variable directly. In the tables that follow in this chapter, we do not show the component variables because they were not included in the final data set.

For some imputed variables in the data set, the number of missing responses does not match the number of imputed responses. Often, the variables correspond to questions that follow a filter question. For example, Item I33 asks if the respondent has difficulty climbing 10 steps; if the response is "yes," the follow-up question (Item I34) asks if the respondent is able to climb 10 steps at all. 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, causing 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

In the following several tables, we present information on how the NBS applied imputation, including the imputed variable names and a brief description of each variable as well as the methods of imputation, total number of missing responses, number of respondents eligible for the question, and percentage of imputed responses. We recorded this information in 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, and (7) longitudinal imputation (using data from an earlier round).[51] In most cases, the logical assignments relied on imputed values. Therefore, the distinction between "logically assigned" and "constructed from other variables with imputed values" is somewhat opaque. In general, if we made a logical assignment for variables corresponding directly to questionnaire questions, we set the flag to 1. For variables constructed from these variables (constructed variables are prefixed with a "C_"), we set the flag to 6. In this instance, we imputed

[51] In prior rounds, the survey had a longitudinal component which Round 4 did not have. Therefore, a longitudinal imputation was considerably more common in prior rounds than in Round 4.

one or more of the component variables in the constructed variable. All variables that include imputed values are identified with the suffix "_i."

In the sections that follow, we summarize the imputations that we conducted, organized by the sections within the questionnaire to which the variables correspond, and provide details for some of the imputation types for each section.

1. Section L: Race and Ethnicity

Several questions gathered information on respondents' race and ethnicity. Two of the variables in Section L included imputed responses, as 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.

Table IV.1. Race and Ethnicity Imputations

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
L1_i	Hispanic/Latino ethnic origins	2 imputations from SSA's administrative data, 3 logical imputations, 108 imputations from hot deck	113	5,078	2.23
C_Race_i	Race	46 imputations from SSA's administrative data, 222 imputations from hot deck	268	5,078	5.28

Source: NBS, Round 4.

In the above table, respondents who did not indicate in the questionnaire whether they were Hispanic were classified as such if the SSA administrative data so indicated; we conducted the single logical imputation 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, we imputed the Hispanic indicator by using a hot deck with imputation classes defined by the ZIP code of each sample member, with race as a sorting variable. Not surprisingly, the imputation classes based on ZIP code commonly required collapsing to ensure that an imputation class had a sufficient number of donors for the recipients in that class. An automated process in SAS performed the needed check. However, to ensure that the ZIP code imputation classes being collapsed were as similar as possible, we manipulated the software so that the county of the donor ZIP code and county of the recipient ZIP code had a similar racial/ethnic composition according to data from the Area Resource File, a file with demographic, health, and economic-related data for every county in the United States (Area Resource File 2009–2010).

Respondents could choose from five race categories—white, black/African American, Asian, Hawaiian/Pacific Islander, and Native American/American Indian—and could select more than one of the categories to identify themselves (as prescribed by the Office of Management and Budget). The final race variable on which imputation was applied included 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 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, resulting in the misclassification of respondents—with SSA administrative data—who did not answer the race question in the survey but would have identified themselves as multiple race or Hawaiian/Pacific Islander. However, we assumed that the number of such respondents was small and that their misclassification was not a major problem. As with the Hispanic indicator, for respondents still with missing data, we imputed race by using a hot deck with imputation classes defined by the ZIP code of each sample member, with Hispanicity as a sorting variable. In general, if the respondent was a longitudinal case then we used the imputed value from earlier rounds. However, the absence of a longitudinal component made longitudinal imputations very rare in Round 4. We did not impute any cases for the race and ethnicity variables using data from earlier rounds.

2. Section B: Disability Status Variables and Work Indicator

In Table IV.2, we describe five imputed variables that pertain to the sample member's disability status and an indicator of whether the respondent was currently working. The imputed variables include three that collapse and recode primary diagnosis codes from the ICD-9 in three ways: C_MainConBodyGroup_i, which corresponds to the collapsing 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). We also imputed a fourth variable with collapsed primary diagnosis codes, with levels further collapsed from C_MainConDiagGrp_i. Table IV.2 does not include this variable (C_MainConImput_i) because it was not released to the final file but 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. Despite the absence of a longitudinal component in Round 4, a few cases selected for Round 4 were part of the sample for one or more of the earlier rounds. For two missing values among these cases, we obtained the age variable from earlier rounds, one from Round 1 and one from Round 3. All missing values for C_AdultChildOnset_i were "logically assigned" by 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 (Item B24_i) in what is a gate question for all of Section C's work status variables.

Table IV.2. Disability Status Imputations

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
C_MainConDiagGrp_i	Primary diagnosis group	84 hot deck	84	4,540	1.85
C_MainConColDiagGrp_i	Main condition diagnosis group collapsed	84 constructed from imputed variables	84	4,540	1.85
C_MainConBodyGroup_i	Main condition body group	8 hot deck, 76 constructed from imputed variables	84	4,540	1.85
C_Disage_i	Age at onset of disability	175 hot deck, 2 from longitudinal data	177	5,078	3.49
C_Adultchild_onset_i	Adult/child onset of disability	21 constructed from imputed variables	21	5,078	0.41
B24_i	Currently working	4 hot deck	4	5,078	0.08

Source: NBS, Round 4.

To define imputation classes, all of the variables in Section B used an indicator to specify whether the onset of the disability occurred in childhood or adulthood and to specify age and gender. We also used one of the collapsed condition code variables, C_MainConImput_i, as a classing variable for disability age and the work indicator. We used additional classing variables specific to the variable being imputed.

3. Section C: Current Jobs Variables

Several survey questions asked respondents about current employment. Section C asked such questions only of respondents who indicated in Item B24 that they were currently working; as identified in Table IV.3, the questions asked about 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).

We imputed values for other variables 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) for the job was missing, we used the relationship between gross monthly and take-home monthly pay among respondents missing neither variable to determine the appropriate value for gross monthly pay. In particular, a random draw was selected from the observed distribution of relative taxes, where "relative tax" is defined as the proportion of a respondent's pay devoted to tax. We then used the randomly drawn relative tax to determine an imputed gross monthly pay for 11 cases with missing data for C_MainCurJobMnthPay. As noted in Table IV.3, we applied hot-deck imputations to only four of the jobs variables: C1_i, C2_1_1d_i, C8_1_i, and C_TotCurMnthPay_i. For these variables, we used the level of education as a classing variable as well as additional classing and sorting variables specific to each variable, including a condition code variable for all but C_TotCurMnthPay_i.

Table IV.3. Current Jobs Imputations

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
C1_i	Count of current jobs	1 hot deck	1	1,023	0.10
C2_1_1d_i	Main current job SOC code to one digit	4 hot deck[a]	4	1,023	0.39
C8_1_i	Hours per week usually worked at current main job	31 hot deck,[b] 2 imputed by distributional assumptions	33	1,023	3.23
C_TotCurWkHrs_i	Total weekly hours at all current jobs	31 hot deck,[c] 4 constructed from imputed variables	35	1,023	3.42
C_TotCurHrMnth_i	Total hours per month at all current jobs	35 constructed from imputed variables	35	1,023	3.42
C_MainCurJobHrPay_i	Hourly pay at current main job	4 logical, 112 constructed from imputed variables	116	1,023	11.34

Table IV.3 *(continued)*

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
C_MainCurJobMnthPay_i	Monthly pay at current main job	21 logical, 10 imputed by distributional assumptions, 102 constructed from imputed variables	133	1,023	13.00
C_TotCurMnthPay_i	Total monthly salary all current jobs	29 logical, 102 hot deck, 6 constructed from imputed variables	137	1,023	13.39

Source: NBS, Round 4.

[a] Imputations for current job variables excluded four cases coded as "don't know" or "refused" in Item B24, which were imputed as currently not working in Item B24_i.

[b] Imputations for current job variables excluded four cases coded as "don't know" or "refused" in Item B24, which were imputed as currently not working in Item B24_i.

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

Some of the variables in the above table had missing values that were not directly imputed. Rather, constituent variables not included in the table had missing values that were imputed and then combined to form the variables in the table. For example, we constructed C_TotCurWkHrs_i 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 we set C_TotCurWkHrs_i equal to C8_1_i. However, if the respondent worked more than one job and the number of hours in secondary jobs was imputed, we constructed C_TotCurWkHrs_i from imputed variables.

4. Section I: Health Status Variables

Section I of the NBS accounts for 56 health status variables where imputations were applied. Tables IV.4 and IV.5 identify the 56 imputed variables and the methods of imputation used for each variable. The items cover a range of topics, from the respondent's general health to specific questions on instrumental activities of daily living (IADLs) and activities of daily living (ADLs) and other health and coping indicators. Included, too, in Section I is a series of questions pertaining to the respondent's use of illicit drugs and alcohol.

Table IV.4. Health Status Imputations, Questionnaire Variables

Variable Name	Description	Imputation Method(s)	Number Missing	Number Eligible	Percent Imputed
I1_i	Health during the past four weeks	10 hot deck	10	5,078	0.20
I9_i	Current health	30 hot deck	30	5,078	0.59
I17a_i	Wears glasses	19 hot deck	19	5,078	0.37
I17b_i	Difficulty seeing with glasses	9 logical, 33 hot deck	42	3,422	1.23
I18_i	Difficulty seeing no glasses	42 logical, 19 hot deck	59	1,698	3.47
I19_i	Uses special equipment because of difficulty seeing	38 logical, 11 hot deck	49	2,113	2.32
I21_i	Difficulty hearing	1 logical, 34 hot deck	35	5,078	0.69
I22_i	Able to hear normal conversation	29 logical, 8 hot deck	37	953	3.88
I23_i	Uses special equipment because of difficulty hearing	29 logical, 3 hot deck	32	953	3.36
I25_i	Difficulty having speech understood	4 logical, 31 hot deck	35	5,078	0.69
I26_i	Able to have speech understood at all	27 logical, 15 hot deck	42	1,339	3.14
I27_i	Uses special equipment because of difficulty speaking	27 logical, 5 hot deck	32	1,339	2.39
I29_i	Difficulty walking without assistance	14 logical, 36 hot deck	50	5,078	0.98
I30_i	Able to walk ¼ mile	21 logical, 56 hot deck	77	2,170	3.55
I31_i	Uses special equipment because of difficulty walking	21 logical, 13 hot deck	34	2,170	1.57
I33_i	Difficulty climbing 10 steps	1 logical, 51 hot deck	52	5,078	1.02
I34_i	Able to climb 10 steps at all	33 logical, 25 hot deck	58	2,210	2.62
I35_i	Difficulty lifting and carrying 10 pounds	4 logical, 35 hot deck	39	5,078	0.77
I36_i	Able to lift or carry 10 pounds at all	23 logical, 27 hot deck	50	2,053	2.44
I37_i	Difficulty using hands or fingers	1 logical, 33 hot deck	34	5,078	0.67

Table IV.4 *(continued)*

Variable Name	Description	Imputation Method(s)	Number Missing	Number Eligible	Percent Imputed
I38_i	Able to use hands or fingers at all	26 logical, 16 hot deck	42	1,157	3.63
I39_i	Difficulty reaching over head	2 logical, 39 hot deck	41	5,078	0.81
I40_i	Able to reach over head at all	22 logical, 20 hot deck	42	1,218	3.45
I41_i	Difficulty standing	52 hot deck	52	5,078	1.02
I42_i	Able to stand at all	25 logical, 13 hot deck	38	2,812	1.35
I43_i	Difficulty stooping	1 logical, 38 hot deck	39	5,078	0.77
I44_i	Able to stoop at all	18 logical, 43 hot deck	61	2,794	2.18
I45_i	Difficulty getting around inside home	1 logical, 26 hot deck	27	5,078	0.53
I46_i	Needs help to get around inside home	24 logical, 5 hot deck	29	769	3.77
I47_i	Difficulty getting around inside home	6 logical, 40 hot deck	46	5,078	0.91
I48_i	Needs help to get around outside home	24 logical, 21 hot deck	45	1,809	2.49
I49_i	Difficulty getting into/out of bed	2 logical, 39 hot deck	41	5,078	0.81
I50_i	Needs help getting into/out of bed	30 logical, 17 hot deck	47	1,309	3.59
I51_i	Difficulty bathing or dressing	4 logical, 32 hot deck	36	5,078	0.71
I52_i	Needs help bathing or dressing	27 logical, 10 hot deck	37	1,031	3.59
I53_i	Difficulty shopping	18 logical, 29 hot deck	47	5,078	0.93
I54_i	Needs help shopping	20 logical, 10 hot deck	30	1,463	2.05
I55_i	Difficulty preparing own meals	7 logical, 28 hot deck	35	5,078	0.69
I56_i	Needs help to prepare meals	22 logical, 12 hot deck	34	1,530	2.22
I57_i	Difficulty eating	25 hot deck	25	5,078	0.49
I58_i	Needs help to eat	23 logical, 4 hot deck	27	638	4.23
I59_i	Trouble concentrating	58 hot deck	58	5,078	1.14

Table IV.4 *(continued)*

Variable Name	Description	Imputation Method(s)	Number Missing	Number Eligible	Percent Imputed
I60_i	Trouble coping with stress	63 hot deck	63	5,078	1.24
I61_i	Trouble getting along with people	73 hot deck	73	5,078	1.44
CageScore_indicator_i	CAGE Alcohol Score	31 constructed from imputed variables	31	4,960	0.63
I72_i	Uses drugs in larger amounts than prescribed	46 hot deck	46	5,078	0.91

Source: NBS, Round 4.

Table IV.5. Health Status Imputations, Constructed Variables

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
C_EquipFuncLim_I	Uses equipment/device for functional/sensory limitation	23 constructed from imputed variables	23	5,078	0.45
C_NumSenLim_i	Number of sensory limitations	85 constructed from imputed variables	85	5,078	1.67
C_NumSevSenLim_i	Number of severe sensory limitations	41 constructed from imputed variables	41	5,078	0.80
C_NumPhyLim_i	Number of physical functional limitations	144 constructed from imputed variables	144	5,078	2.84
C_NumSevPhyLim_i	Number of severe physical functional limitations	168 constructed from imputed variables	168	5,078	3.31
C_NumEmotLim_i	Number of emotional/social limitations	125 constructed from imputed variables	125	5,078	2.46
C_NumADLs_i	Number of impaired ADL	56 constructed from imputed variables	56	5,078	1.10
C_NumADLAssist_i	Number of ADL requiring assistance	54 constructed from imputed variables	54	5,078	1.06
C_NumIADLs_i	Number of IADL difficulties	68 constructed from imputed variables	68	5,078	1.34
C_NumIADLAssist_i	Number of IADL requiring assistance	37 constructed from imputed variables	37	5,078	0.73
C_PCS8TOT_i	Physical summary score	148 constructed from imputed variables	148	5,078	2.91

Table IV.5 *(continued)*

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
C_MCS8TOT_i	Mental summary score	148 constructed from imputed variables	148	5,078	2.91
C_DrugDep_i	Drug dependence	47 constructed from imputed variables	47	5,078	0.93

Source: NBS, Round 4.

The following is an example of a logical assignment in Section I: if a respondent did not answer whether he or she experienced difficulty in seeing newsprint letters even when wearing glasses or contact lenses (Item I17b) but indicated that he or she could not see newsprint letters at all (Item I18) or required special devices to read newsprint letters (Item I19), then we logically assigned "yes" to Item I17b_i.

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

The only classing variable common to all imputations was the collapsed condition code variable. We also used age and gender in most imputations. The other classing and sorting variables were specific to the variable being imputed.

5. Section K: Sources of Income Other than Employment

The imputed variables in Section K are constructed variables that pertain to nonemployment-based income and include workers' compensation, private disability claims, unemployment, and other sources of regular income, as described in Table IV.6.

Table IV.6. Imputations on Sources of Income Other than Employment

Variable Name	Description	Imputation Methods	Number Missing	Number Eligible	Percent Imputed
C_AmtPrivDis_i	Amount received from private disability last month	90 logical, 16 imputed by descriptive statistics using specialized procedures	107	5,078	2.11
C_AmtWorkComp_i	Amount received from workers' compensation last month	51 logical, 8 imputed by descriptive statistics using specialized procedures	59	5,078	1.16
C_AmtVetBen_i	Amount received from veterans' benefits last month	45 logical, 9 imputed by descriptive statistics using specialized procedures	54	5,078	1.06
C_AmtPubAssis_i	Amount received from public assistance last month	65 logical, 25 imputed by descriptive statistics using specialized procedures	90	5,078	1.77

Table IV.6 *(continued)*

Variable Name	Description	Imputation Methods	Number Missing	Number Eligible	Percent Imputed
C_AmtUnemply_i	Amount received from unemployment benefits last month	48 logical, 2 imputed by descriptive statistics using specialized procedures	50	5,078	0.98
C_AmtPrivPen_i	Amount received from private pension last month	50 logical, 17 imputed by descriptive statistics using specialized procedures	67	5,078	1.32
C_AmtOthReg_i	Amount received from other regular sources last month	44 logical, 20 imputed by descriptive statistics using specialized procedures	64	5,078	1.26

Source: NBS, Round 4.

Items in Section K first asked respondents if they received money from a specific source and then asked for the specific amount received from that source. If a respondent could not provide a specific value, he or she either answered a series of questions about whether the amount was above or below specific values or had the option of providing a range of values, where the options depended on responses to a series of questions. After we classified the response according to a range of values provided by the respondent, we assigned the respondent the median of the specific values provided by others 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 first imputed the range (using random assignment) and then assigned the median of the values provided by respondents who listed specific values within that range. If the respondent did not know if he or she received funds from a source, we used hot-deck imputation to determine whether such was the case and then proceeded as above.

The logical assignments in Section K derive from imputed values in the constituent questions. For example, K6 in the questionnaire asks whether the respondent received income from a variety of sources, and K7 asks the amount from each source for which a "yes" response was given. The first source listed (K6a) is private disability insurance. If the respondent was imputed not to have received private disability insurance (K6a_i), then the constructed variable C_AmtPrivDis_i (based on K7) was logically assigned "no." Otherwise, if any income was derived from private disability insurance, but an imputation was required at some point in the sequence (either everything or just the individual's income was imputed), then the imputation flag indicated imputation by "special procedures."

For variables requiring hot-deck imputation, 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. Table IV.6 lists none of the variables requiring hot-deck imputation because they were just 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. The questions from which the imputed variables were derived ask about education (L3_i), marital status (L8_i), cohabitation status

(C_Cohab_i), number of children in household (C_NumChildHH_i), household size (C_Hhsize_i), and weight and height, which were used to derive body mass index (C_BMI_cat_i). Most of these variables were imputed early in imputation processing and were used in the imputation of variables imputed later in processing.[52] Household income questions are also asked in Section L, which, in combination with C_Hhsize_i and C_Numchildhh_i, we use to derive the Federal poverty level variable.

The imputation of poverty level required the imputation of annual income and household size. The annual income question was another case that required a specific value; if the respondent could not provide a specific value, he or she was asked if annual income fell within certain ranges. Some respondents provided a specific value, some provided a range of values, 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 Table IV.7 because it was not released in the final file. All missing values in C_FedPovertyLevel_cat[53] were derived from the imputed annual incomes; hence, all missing values are "constructed from imputed variables." In Table IV.7, we identify the imputed variables in Section L.

Logical assignments in Section L are based on related variables also in Section L. For example, the two logical assignments for L11_i (living situation of beneficiary) are attributable to the fact that two respondents did not answer L11 but indicated in L16 (number of adults in household) that only one adult lived in the household and indicated in L17 (number in household under 18 years old) the number of children living in the household. For these two respondents, the value for L11_i was logically assigned to 1 or 2, depending on the response to L17.

The only classing variable common to all imputations for the variables listed in Table IV.7 was the collapsed condition code variable. Other classing and sorting variables were specific to the variable being imputed.

52 An additional variable C_NumChildren_i was also imputed. It is defined as the total number of children in the household plus the number of respondent's children living outside the household. None of the subsequent processing used this variable, which, on further review, was not deemed necessary for analysis, although it is in the final file.

53 The name of this variable reflects that fact that the final variable was a categorical (as opposed to a continuous) measure of poverty level.

Table IV.7. Imputations of Personal and Household Characteristics

Variable Name	Description	Imputation Method(s)	Number Missing	Number Eligible	Percent Imputed
C_BMI_Cat_i	Body Mass Index categories	185 hot deck	185	5,078	3.64
L3_i	Highest year/grade completed in school	99 hot deck	99	5,078	1.95
L8_i	Marital status	53 hot deck	53	5,078	1.04
L11_i	Living arrangements	3 logical, 48 hot deck	51	5,078	1.00
C_NumChildhh_i	Number of children living in household	3 logical, 32 hot deck, 1 constructed from imputed variables	36	5,078	0.71
C_hhsize_i	Household size	57 hot deck, 14 constructed from imputed variables	71	5,078	1.40
C_cohab_i	Cohabitation status	5 logical, 47 hot deck	52	5,078	1.02
C_FedPovertyLevel_cat1	2009 Federal poverty level	1,707 constructed from imputed variables	1,707	5,078	33.62

Source: NBS, Round 4.

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. The latter include the ratio of two estimates (for example, a mean or proportion in which both the numerator and 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 generally are not available, and various approximations or computational algorithms 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 to obtain the correct estimate. Most procedures in standard statistical packages, such as SAS, STATA, 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 likely underestimates substantially the sampling variance with a design as complex as that used for the NBS. 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.[54]

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 that use explicit sampling variance equations and the procedures based on forming pseudo-replications[55] of the sample. The Taylor series linearization procedure is based on a classic statistical method in which a nonlinear statistic may be approximated by a linear combination of the components within the statistic. The accuracy of the approximation depends 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 may be

[54] A 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 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 the 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.

[55] Pseudo-replications of a specific survey sample, as opposed to true replications of the sampling design, involve the selection of several independent subsamples from the original sample data with the same sampling design. The subsamples may be random (as in a bootstrap) or restricted (as in Balanced Repeated Replication).

used to estimate the sampling variance, and the sampling variance may be estimated by using many features of the sampling design (for example, finite population corrections, stratification, multiple stages of selection, and unequal selection rates within strata). This is the basic variance estimation procedure used in all SUDAAN procedures as well as in the survey procedures in SAS, STATA, and other software packages that accommodate simple and complex sampling designs. To calculate the variance, sample design information (such as stratum, analysis weight, and so on) is needed for each sample unit.

Currently, several survey data analysis software packages use the Taylor series linearization procedure and explicit sampling variance equations. Therefore, we developed the variance estimation specifications needed for the Taylor series linearization (PseudoStrata and PseudoPSU). Appendix E provides example code for the procedure with SAS and the survey data analysis software SUDAAN.[56] Details about SAS syntax are available from SAS (SAS Institute 2004); details about SUDAAN syntax are available from RTI International (Research Triangle Institute 2004).

[56] The example code provided in Appendix F 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 complex analyses. Given that SUDAAN was created specifically for survey data, the range of analyses that may be performed with these data in SUDAAN is much wider than that in SAS.

VI. REFERENCES

Agresti, A. Categorical Data Analysis. New York: John Wiley and Sons, 1990.

Akaike, H. "A New Look at the Statistical Model Identification." *IEEE Transaction on Automatic Control*, AC-19, 1974, pp. 716-723.

Area Resource File (ARF). Rockville, MD: U.S. Department of Health and Human Services, Health Resources and Services Administration, Bureau of Health Professions, 2009–2010.

Barrett, K., S. Skidmore, and D. Wright. "National Beneficiary Survey Round 4 (volume 2 of 3): Cleaning and Identification of Data Problems." Washington, DC: Mathematica Policy Research, February 2012.

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.

Grau, E., K. Barrett, D. Wright, Y. Zheng, Barbara Carlson, F. Potter, and S. Skidmore. "National Beneficiary Survey Round 4 (volume 1 of 3): Editing, Coding, Imputation, and Weighting Procedures." Washington, DC: Mathematica Policy Research, February 2012.

Grau, E., K. Barrett, and D. Wright. "National Beneficiary Survey Round 4: Nonresponse Bias Analysis." Washington, DC: Mathematica Policy Research, February 2012.

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.

Hosmer, D.W., Jr., and S. Lemeshow. "Goodness-of-Fit Tests for the Multiple Logistic Regression Model. *Communications in Statistics, Theory and Methods*, vol. A9, no. 10, 1980, pp. 1043-1069.

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.

Rall, K., D. Wright, R. Scurato, and S. Khambhati. "The National Beneficiary Survey: Round 4 Public-Use File Codebook." Washington, DC: Mathematica Policy Research, February 2012.

Rall, K., D. Wright, R. Scurato, and S. Khambhati. "The National Beneficiary Survey: Round 4 Restricted-Use File Codebook." Washington, DC: Mathematica Policy Research, February 2012.

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, 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, 2004.

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

Wright, D., K. Barrett., E. Grau, Y. Zheng, and S. Skidmore. "National Beneficiary Survey Round 4 (volume 3 of 3): User's Guide for Restricted and Public Use Data Files." Washington, DC: Mathematica Policy Research, February 2012.

Wright, D., K. Barrett, S. Skidmore. "National Beneficiary Survey: Round 4 Questionnaire." Washington, DC: Mathematica Policy Research, February 2012.

APPENDIX A

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

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

Appendix A. “Other/Specify” and Open- Ended Items with Additional Categories Created During Coding

Question #	Question Text	Current Response Options	Additional Categories Created
B25	What are they (the other reasons you are not working that I didn’t mention)?	a = A physical or mental condition prevents [you/him/her] from working b = [You/NAME] cannot find a job that [you are/(he/she) is] qualified for c = [You do/NAME does] not have reliable transportation to and from work d = [You are/NAME is] caring for someone else. f = [You/NAME] cannot find a job [you want/(he/she) wants] g = [You are/NAME is] waiting to finish school or a training program. h = Workplaces are not accessible to people with [your/NAME’s] disability. i = [You do/NAME does] not want to lose benefits such as disability, worker’s compensation, or Medicaid j = [Your/NAME’s] previous attempts to work have been discouraging l = Others do not think [you/NAME] can work m = Employers will not give [you/NAME] a chance to show that [you/he/she] can work. n = [You/NAME] does not have the special equipment or medical devices that [you/he/she] would need in order to work. o = [You/NAME] cannot get the personal assistance [you need/he needs/she needs] in order to get ready for work each day	p=Cannot find a job/job market is bad q=Lack skills
B29_6	What benefits [were/was] [you/NAME] most worried about losing?	1= Private disability insurance 2= Workers’ compensation 3= Veterans’ benefits 4= Medicare 5= Medicaid 6= SSA disability benefits 7= Public assistance or welfare 8= Food stamps 9= Personal assistance services (pas) 10= Unemployment benefits 11= Other state disability benefits 12= Other government programs 13= Other	14= Health insurance unspecified

Appendix A *(continued)*

Question #	Question Text	Current Response Options	Additional Categories Created
B29_10	What benefits [were/was] [you/NAME] most worried about losing?	01= Private Disability Insurance 02= Workers' compensation 03= Veterans' benefits 04= Medicare 05= Medicaid 06= SSA Disability Benefits 07= Public Assistance or Welfare 08= Food Stamps 09= Personal Assistance Services (PAS) 10= Unemployment Benefits 11= Other State Disability Benefits 12= Other government programs 13= Other	14= Health insurance unspecified
B29_11b	What benefits [were/was] [you/NAME] most worried about losing?	01= Private Disability Insurance 02= Workers' compensation 03= Veterans' benefits 04= Medicare 05= Medicaid 06= SSA Disability Benefits 07= Public Assistance or Welfare 08= Food Stamps 09= Personal Assistance Services (PAS) 10= Unemployment Benefits 11= Other State Disability Benefits 12= Other government programs 13= Other	14= Health insurance unspecified
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 <u>not</u> 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]?

Appendix A *(continued)*

Question #	Question Text	Current Response Options	Additional Categories Created
C39b	[Do you/Does NAME] work fewer hours or earn less money than [you/he/she] could because [you/he/she]:	a = [Are/Is] taking care of children or others? b = [Are/Is] enrolled in school or a training program? c = Want[s] to keep Medicare or Medicaid coverage? d = Want[s] to keep cash benefits [you/he/she] need such as disability or workers' compensation? e = Just [do/does] not want to work more? f = Are there any reasons I didn't mention why [you are/NAME is] working or earning less than [you/he/she] could?	g=[Are/is] in poor health or [have/has] health concerns?
C39_2	What benefits have been reduced or ended as a result of [your/NAME's] (main/current) job?	01 = Private Disability Insurance 02 = Workers' compensation 03 = Veterans' benefits 04 = Medicare 05 = Medicaid 06 = SSA Disability Benefits 07 = Public Assistance or Welfare 08 = Food Stamps 09 = Personal Assistance Services (PAS) 10 = Unemployment Benefits 11 = Other State Disability Benefits 12 = Other government programs 13 = Other	14= Health insurance unspecified

Appendix A *(continued)*

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
D25	Did you work fewer hours or earn less money than you could have because [you/he/she] you…	a= [Were/Was] taking care of somebody else? b= [Were/Was] enrolled in school or a training program? c= Wanted to keep Medicare or Medicaid coverage d= Wanted to keep cash benefits such as disability or workers compensation? e= Just didn't want to work more? f= Are there any reasons I didn't mention why [you/NAME] might have chosen to work or earn less than [you/he/she] could have during 2004? (SPECIFY: <OPEN>)	g=Had medical problems/complications

Appendix A *(continued)*

Question #	Question Text	Current Response Options	Additional Categories Created
D26	In 2009, 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: <OPEN>) h=Is there anything else that I didn't mention that would have helped [you/NAME] to work or earn more during 2004? (SPECIFY: <OPEN>)	i=Better health/treatment j=More supportive/helpful employer and/or coworker
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
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
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

Appendix A *(continued)*

Question #	Question Text	Current Response Options	Additional Categories Created
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? c=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
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=Other 04=Recommended/told to use it/thought using it was required
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 03=Other

Appendix A *(continued)*

Question #	Question Text	Current Response Options	Additional Categories Created
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=Other problems 05=Transportation/location problems
H48	What was the problem about?	<OPEN>	01=Trouble making/keeping contact 02=Did not receive services wanted/needed 03=Other problems
K14	What other assistance did [you/NAME] receive <u>last month</u>?	<OPEN>	01=Housing Assistance 02=Energy Assistance 03=Food assistance 04=Other

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

APPENDIX B

SOC MAJOR AND MINOR OCCUPATION CLASSIFICATIONS

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

Appendix B. 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

Appendix B *(continued)*

Code	Occupation
	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
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

Appendix B *(continued)*

Code	Occupation
	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
	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

Appendix B *(continued)*

Code	Occupation
	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

APPENDIX C

NAICS INDUSTRY CODES

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

Appendix C. 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

Appendix C *(continued)*

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

Appendix C *(continued)*

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

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

APPENDIX D

PARAMETER ESTIMATES AND STANDARD ERRORS FOR NONRESPONSE MODELS

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

Table D.1. Variables in the Location Logistic Propensity Model Representative Beneficiary Sample

Main Effects	Parameter Estimate[a]	Standard Error
Variables in the Beneficiary Location Model		
Count of addresses on file (MOVE)		
Only one address on file	-0.553	0.620
Two addresses on file	-1.720**	0.630
Three or more addresses on file	-2.517**	0.629
No information	Ref. cell	
Count of phone numbers on file (PHONE)		
One or two phone numbers on file	-0.403	0.401
Three or more phone numbers on file	-1.658**	0.000
No information	Ref. cell	
Geographic region (based on U.S. Census regions) of beneficiary's residence (REGION)		
South	0.646**	0.215
West/Midwest/Northeast	Ref. cell	
Urbanicity of place of residence of beneficiary (METRO)		
Beneficiary resides in metropolitan statistical area (MSA) of 1 million or more	1.029*	0.460
Beneficiary resides in metropolitan statistical area (MSA) of 250,000-999,999	0.979*	0.475
Beneficiary resides in metropolitan statistical area (MSA) of less than 250,000	0.805	0.497
Beneficiary resides in nonmetropolitan area adjacent to large metropolitan area	1.269*	0.583
Beneficiary resides in nonmetropolitan area adjacent to small metropolitan area	1.047	0.619
Beneficiary resides in nonmetropolitan area not adjacent to metropolitan area	Ref. cell	
Race of the beneficiary (RACE)		
White	0.856†	0.231
Not white or unknown	Ref. cell	
Racial/ethnic makeup of county (CNTYRACE)		
County with racially/ethnically mixed population, no majority group	0.592†	0.275
County that doesn't have this attribute	Ref. cell	
Population loss county (CNTYPOPLOSS)		
County that experienced population loss 1980-1990 & 1990-2000	0.744*	0.312
County that doesn't have this attribute	Ref. cell	
Low education county (CNTYLOWEDUC)		
County with low levels of education	0.504	0.267
County that doesn't have this attribute	Ref. cell	
Two- Factor Interactions[b]		
RACE*CNTYRACE		
Racially/ethnically mixed county*White	-1.824**	0.394

[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% and 1% levels respectively.

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

Table D.2. Variables in the Cooperation Logistic Propensity Model Representative Beneficiary Sample

Main Effects	Parameter Estimate[a]	Standard Error
Variables in the Beneficiary Cooperation Model		
Beneficiary's age category (AGECAT)		
Age in range 18 to 29 years	-0.255†	0.145
Age in range 30 to 39 years	-0.252†	0.138
Age in range 40 to 49 years	-0.208†	0.137
Age in range 50 to 64 years	Ref. cell	
Race of the beneficiary (RACE)		
Black	0.397**	0.150
Not Black or Unknown	Ref. cell	
Whether the beneficiary was Hispanic or not (HISPANICITY)		
Hispanic	-0.906†	0.805
Nonhispanic/Unknown	Ref. cell	
Urbanicity of place of residence of beneficiary (METRO)		
Beneficiary resides in metropolitan statistical area (MSA) of 1 million or more	-0.127†	0.446
Beneficiary resides in metropolitan statistical area (MSA) of 250,000-999,999	0.162†	0.465
Beneficiary resides in metropolitan statistical area (MSA) of less than 250,000	-0.838†	0.523
Beneficiary resides in nonmetropolitan area adjacent to large metropolitan area	-0.889†	0.558
Beneficiary resides in nonmetropolitan area adjacent to medium or small metropolitan area	0.518†	0.559
Beneficiary resides in nonmetropolitan area not adjacent to metropolitan area	Ref. cell	
Geographic region (based on U.S. Census divisions) of beneficiary's place of residence (DIVISION)		
New England	-0.568	0.294
West South Central	-0.674**	0.234
All other Census divisions	Ref. cell	
Beneficiary's gender (GENDER)		
Male	-1.648†	0.538
Female	Ref. cell	
Identity of payee relative to beneficiary (REPREPAYEE)		
Beneficiary received payments himself/herself	-0.842**	0.314
Beneficiary did not receive payments himself/herself, or unknown	Ref. cell	
Indicator whether beneficiary and applicant for benefits are in same zip code (PDZIPSAME)		
Applicant and beneficiary live in same zip code	1.829†	0.500
Applicant and beneficiary live in different zip code, or no information	Ref. cell	
Count of addresses on file (MOVE)		
Only one address on file	-0.070†	0.195
Two addresses on file	0.425†	0.214
Three or more addresses on file	0.552†	0.265
No information	Ref. cell	
Count of phone numbers on file (PHONE)		
One or two phone numbers on file	0.821†	0.325
Three or more phone numbers on file	0.299†	0.227
Unknown	Ref. cell	
Beneficiary's living situation (LIVING)		
Beneficiary lives in an institution	-0.665	0.470
Beneficiary lives with others, or no information	Ref. cell	

Table D.2 *(continued)*

Main Effects	Parameter Estimate[a]	Standard Error
Racial/ethnic makeup of county (CNTYRACE)		
County with plurality or majority Hispanic population	-0.323	0.228
County with a racially/ethnically mixed population, no majority group	-0.311	0.148
County that doesn't have these attributes	Ref. cell	
Government- dependent county (CNTYGOV)		
County with a government-dependent economy	-0.376	0.207
County that doesn't have this attribute	Ref. cell	
Two- Factor Interactions[b]		
PDZIPSAME*PHONE		
Applicant & beneficiary live in same zip code*One or two phone numbers on file	-1.047**	0.396
Applicant & beneficiary live in same zip code*Three or more phone numbers on file	-0.206	0.257
PDZIPSAME*METRO		
Applicant & beneficiary live in same zip code*Metropolitan areas 1 million or more	-1.282*	0.526
Applicant & beneficiary live in same zip code*Metropolitan areas 250,000-999,999	-1.530**	0.542
Applicant & beneficiary live in same zip code*Metropolitan areas less than 250,000	-1.327*	0.607
Applicant & beneficiary live in same zip code*Nonmetropolitan areas adjacent to large metropolitan area	-0.573	0.648
Applicant & beneficiary live in same zip code*Nonmetropolitan areas adjacent to medium or small metropolitan areas	-1.308	0.702
GENDER*METRO		
Male*Metropolitan areas 1 million or more	1.326*	0.569
Male*Metropolitan areas 250,000-999,999	1.352*	0.589
Male*Metropolitan areas less than 250,000	2.708**	0.641
Male*Nonmetropolitan areas adjacent to large metropolitan area	2.306**	0.696
Male*Nonmetropolitan areas adjacent to medium or small metropolitan areas	1.246	0.722
HISPANIC*AGECAT		
Hispanic*Age 18 to 29 years	1.454**	0.561
Hispanic*Age 30 to 39 years	1.084	0.606
Hispanic*Age 40 to 49 years	0.673	0.562
HISPANIC*MOVE		
Hispanic*Only one address on file	0.985	0.837
Hispanic*Two addresses on file	-0.496	0.826
Hispanic*Three or more addresses on file	-1.247	0.842

[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% and 1% levels respectively.

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

Table D.3. Variables in the Cooperation Logistic Propensity Model Ticket Participant Sample, State Vocational Rehabilitation Agencies (SVRAS) Acting As Employment Networks (ENS)

Main Effects	Parameter Estimate[a]	Standard Error
Variables in the Location Model for SVRAs Acting as ENs Provider Type		
Count of addresses on file (MOVE)		
Two addresses on file	-0.687*	0.303
Three or more addresses on file	-1.473**	0.333
Only one address on file or No information	Ref. cell	
Count of phone numbers on file (PHONE)		
One or more phone numbers on file	-1.392**	0.391
No information	Ref. cell	
Urbanicity of place of residence of participant (METRO)		
Participant resides in metropolitan statistical area (MSA) of 1 million or more	-1.858**	0.469
Participant resides in metropolitan statistical area (MSA) of 250,000-999,999	-1.319**	0.303
Participant resides in metropolitan areas of less than 250,000 or in nonmetropolitan area	Ref. cell	
Living situation of participant (LIVING)		
Participant lives alone	-0.547*	0.244
Participant does not live alone, or information unknown	Ref. cell	
Racial/ethnic makeup of county (CNTYRACE)		
County with racially/ethnically mixed population, no majority group	-2.153**	0.594
County with at least 90% non-Hispanic white population	-1.817**	0.369
County that doesn't have these attributes	Ref. cell	
Service- dependent economy county (CNTYSVC)		
County with economy dependent upon services	1.948**	0.725
County that doesn't have this attribute	Ref. cell	

[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% and 1% levels respectively.

Table D.4. Variables in the Cooperation Logistic Propensity Model Ticket Participant Sample, State Vocational Rehabilitation Agencies (SVRAS) Acting As Employment Networks (ENS)

Main Effects	Parameter Estimate[a]	Standard Error
Variables in the Cooperation Model for SVRAs Acting as ENs Provider Type		
Participant's age category (AGECAT)		
Age in range 18 to 29 years	0.086†	0.447
Age in range 30 to 39 years	0.700†	0.539
Age in range 40 to 49 years	0.573†	0.362
Age in range 50 to 64 years	Ref. cell	
Identity of payee relative to participant (REPREPAYEE)		
A family member received benefits on behalf of participant	0.336	0.229
Participant received benefit payments himself/herself, an institution received payments on behalf of participant, or information unknown	Ref. cell	
Count of addresses on file (MOVE)		
One or two addresses on file	0.181	0.303
Three or more addresses on file	-0.391	0.384
No information	Ref. cell	
Count of phone numbers on file (MOVE)		
One or two phone numbers on file	-0.726†	0.339
Three or more phone numbers on file	-0.831†	0.308
No information	Ref. cell	
Participant recipient benefit type (SSI_SSDI)		
SSDI only	-0.247	0.210
SSI only, or both SSI and SSDI	Ref. cell	
Geographic region (based on U.S. Census divisions) of participant's place of residence (DIVISION)		
New England	-0.417	0.261
All other Census divisions	Ref. cell	
Disability diagnosis classification (DIG)		
Participant was deaf	-1.553†	0.553
Participant had mental disability	0.840†	0.334
Participant had other physical disability, or information about disability not given	Ref. cell	
Race of the participant (RACE)		
Black	-0.195†	0.310
Race known not to be black, or unknown	Ref. cell	
Gender of participant (GENDER)		
Male	-0.113†	0.278
Female	Ref. cell	
Service- dependent economy county (CNTYSVC)		
County with economy dependent upon services	-0.569*	0.235
County that doesn't have this attribute	Ref. cell	
Nonspecialized- dependent economy county (CNTYNONSP)		
County that did not meet economic thresholds for government-dependent economy, mining-dependent economy, manufacturing-dependent economy, farming-dependent economy, or services-dependent economy	0.871†	0.460
County that doesn't have this attribute	Ref. cell	

Table D.4 *(continued)*

Main Effects	Parameter Estimate[a]	Standard Error
Two- Factor Interactions[b]		
AGECAT*CNTYNONSP		
Age in range 18 to 29 years * County that did not meet economic thresholds for government-dependent economy, mining-dependent economy, manufacturing-dependent economy, farming-dependent economy, or services-dependent economy	-1.398**	0.478
Age in range 30 to 39 years * County that did not meet economic thresholds for government-dependent economy, mining-dependent economy, manufacturing-dependent economy, farming-dependent economy, or services-dependent economy	-1.100	0.648
Age in range 40 to 49 years * County that did not meet economic thresholds for government-dependent economy, mining-dependent economy, manufacturing-dependent economy, farming-dependent economy, or services-dependent economy	-1.521**	0.538
RACE*GENDER		
Black*Male	1.199**	0.460
DIG*AGECAT		
Participant had mental disability * Age in range 18 to 29 years	-0.697	0.519
Participant had mental disability * Age in range 30 to 39 years	-1.544**	0.547
Participant had mental disability * Age in range 40 to 49 years	-0.761	0.503
GENDER*PHONE		
Male*One or two phone numbers on file	0.901	0.512
Male*Three or more phone numbers on file	0.992*	0.483

[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% and 1% levels respectively.

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

Table D.5. Variables Used in the Location Logistic Propensity Model Ticket Participant Sample, Employment Networks That Are Not State Vocational Rehabilitation Agencies (NONSVRA ENS)

Main Effects	Parameter Estimate[a]	Standard Error
Variables in the Location Model for nonSVRAs ENs Provider Type		
Count of addresses on file (MOVE)		
Two addresses on file	0.061†	0.395
Three or more addresses on file	-0.377†	0.432
Only one address on file or No information	Ref. cell	
Count of phone numbers on file (PHONE)		
One to four phone numbers on file	-2.092†	0.544
Five or more phone numbers on file	-2.804†	0.528
No information	Ref. cell	
Urbanicity of place of residence of participant (METRO)		
Participant resides in metropolitan statistical area (MSA) of 1 million or more	0.403	0.305
Participant does not reside in MSA of 1 million or more	Ref. cell	
Living situation of participant (LIVING)		
Participant lives with others	-0.569*	0.373
Participant lives in another living situation, or information unknown	Ref. cell	
Participant recipient benefit type (SSI_SSDI)		
SSI only	-0.291	0.216
SSDI only, or both SSI and SSDI	Ref. cell	
Geographic region (based on U.S. Census divisions) of participant's place of residence (DIVISION)		
New England	-0.573*	0.283
Pacific	-0.506	0.306
All other Census divisions	Ref. cell	
Disability diagnosis classification (DIG)		
Participant had mental disability	-0.900†	0.545
Participant had physical disability, or information unknown	Ref. cell	
Racial/ethnic makeup of county (CNTYRACE)		
County with racially/ethnically mixed population, no majority group	-1.479†	0.586
County that doesn't have this attribute	Ref. cell	
Service- dependent economy county (CNTYSVC)		
County with economy dependent upon services	-0.683*	0.275
County that doesn't have this attribute	Ref. cell	
Two- Factor Interactions[b]		
PHONE*CNTYRACE		
One to four phone numbers on file*County with racially/ethnically mixed population	1.324*	0.586
Five or more phone numbers on file*County with racially/ethnically mixed population	1.138	0.592
PHONE*DIG		
One to four phone numbers on file*Participant had mental disability	1.509*	0.616
Five or more phone numbers on file*County with racially/ethnically mixed population	1.275*	0.606
MOVE*DIG		
Two addresses on file*Participant had mental disability	-0.850	0.554
Three or more addresses on file*County with racially/ethnically mixed population	-1.714**	0.599

Table D.5 *(continued)*

Main Effects	Parameter Estimate[a]	Standard Error
DIG*CNTYRACE		
Participant had mental disability*County with racially/ethnically mixed population	0.866	0.476

[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% and 1% levels respectively.

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

Table D.6. Variables in the Cooperation Logistic Propensity Model Ticket Participant Sample, Employment Networks That Are Not State Vocational Rehabilitation Agencies (NONSVRA ENS)

Main Effects	Parameter Estimate[a]	Standard Error
Variables in the Cooperation Model for nonSVRA ENs Provider Type		
Race of the participant (RACE)		
Black	-0.433†	0.267
Race known to be neither white nor black, or unknown	Ref. cell	
Hispanicity of participant (HISPANIC)		
Hispanic	0.813	0.420
Not Hispanic, or unknown	Ref. cell	
Disability diagnosis classification (DIG)		
Participant was deaf	-1.213†	0.657
Participant had physical disability other than deafness	-0.027†	0.151
Participant had mental disability, or information unknown	Ref. cell	
Indicator whether participant and applicant for benefits are in same zip code (PDZIPSAME)		
Applicant and participant live in same zip code	-0.334†	0.153
Applicant and participant live in different zip code/No information	Ref. cell	
Identity of payee relative to participant (REPREPAYEE)		
A family member received benefits on behalf of participant	-0.300	0.175
Participant received benefit payments himself/herself, an institution received payments on behalf of participant, or information unknown	Ref. cell	
Geographic region (based on U.S. Census regions) of participant's place of residence (REGION)		
Midwest	0.538†	0.461
Northeast, South, West	Ref. cell	
Participant's gender (GENDER)		
Male	-0.070†	0.172
Female	Ref. cell	
Participant recipient benefit type (SSI_SSDI)		
SSDI only	-0.211†	0.178
Both SSI and SSDI	-0.457†	0.230
SSI only	Ref. cell	
Count of addresses on file (MOVE)		
One address on file	-0.260	0.297
Two addresses on file	-0.638*	0.304
Three or more addresses on file	-0.856*	0.385
No information	Ref. cell	
Count of phone numbers on file (PHONE)		
One or two phone numbers on file	-0.541†	0.321
Three or more phone numbers on file	-0.592†	0.269
No information	Ref. cell	
Urbanicity of place of residence of beneficiary (METRO)		
Participant resides in metropolitan statistical area (MSA) of 1 million or more residents	0.025†	0.321
Participant resides in metropolitan area of less than 1 million residents	-0.265†	0.306
Participant resides in nonmetropolitan area	Ref. cell	

Table D.6 *(continued)*

Main Effects	Parameter Estimate[a]	Standard Error
Racial/ethnic makeup of county (CNTYRACE)		
County with racially/ethnically mixed population, no majority group	-0.096	0.204
County with majority Hispanic population	-0.408	0.258
County with majority but less than 90% non-Hispanic white population	-0.299†	0.275
County that doesn't have this attribute	Ref. cell	
County with housing stress (CNTYHOUSSTRESS)		
County with issues related to housing	-0.243†	0.200
County that doesn't have this attribute	Ref. cell	
Government- dependent economy county (CNTYGOV)		
County with economy dependent upon government	0.429	0.282
County that doesn't have this attribute	Ref. cell	
Service- dependent economy county (CNTYSVC)		
County with economy dependent upon services	-0.106†	0.270
County that doesn't have this attribute	Ref. cell	
Nonspecialized- dependent economy county (CNTYNONSP)		
County that did not meet economic thresholds for government-dependent economy, mining-dependent economy, manufacturing-dependent economy, farming-dependent economy, or services-dependent economy	1.156†	0.477
County that doesn't have this attribute	Ref. cell	
Two- Factor Interactions[b]		
PDZIPSAME*CNTYNONSP		
Applicant and participant live in same zip code* County that did not meet economic thresholds for government-dependent economy, mining-dependent economy, manufacturing-dependent economy, farming-dependent economy, or services-dependent economy	-1.569**	0.416
PHONE*CNTYNONSP		
One or two phone numbers on file*County that did not meet economic thresholds for government-dependent economy, mining-dependent economy, manufacturing-dependent economy, farming-dependent economy, or services-dependent economy	-0.251**	0.512
Three or more phone numbers on file*County that did not meet economic thresholds for government-dependent economy, mining-dependent economy, manufacturing-dependent economy, farming-dependent economy, or services-dependent economy	0.685**	0.411
PHONE*GENDER		
Male*One or two phone numbers on file	0.643	0.373
Male*Three or more phone numbers on file	-0.260	0.288
REGION*METRO		
Midwest*Participant lives in metropolitan areas of 1 million or more	-0.665	0.515
Midwest*Participant lives in metropolitan areas of under 1 million	0.372	0.572
RACE*PHONE		
Black*One or two phone numbers on file	-0.301	0.395
Black*Three or more phone numbers on file	0.863**	0.280
SSI_SSDI*PHONE		
Both SSI & SSDI*One or two phone numbers on file	0.907*	0.401
Both SSI & SSDI*Three or more phone numbers on file	0.183	0.323

Table D.6 *(continued)*

Main Effects	Parameter Estimate[a]	Standard Error
DIG*CNTYNONSP		
Participant had physical disability*County that did not meet economic thresholds for government-dependent economy, mining-dependent economy, manufacturing-dependent economy, farming-dependent economy, or services-dependent economy	0.779*	0.388
RACE*CNTYHOUSSTRESS		
Black*County with issues related to housing	0.618*	0.295
CNTYRACE*CNTYSVC		
County with majority but less than 90% non-Hispanic white population*County with economy dependent upon services	0.548	0.326

[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% and 1% levels respectively.

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

Table D.7. Variables in the Location Logistic Propensity Model Ticket Participant Sample, Traditional Payment System

Main Effects	Parameter Estimate[a]	Standard Error
Variables in the Location Model for Participants Using the Traditional Payment System		
Participant's age category (AGECAT)		
Age in range 18 to 29 years	-1.059*	0.433
Age in range 30 to 39 years	-0.552	0.386
Age in range 40 to 49 years	-0.373	0.466
Age in range 50 to 64 years	Ref. cell	
Race of the participant (RACE)		
Black	0.712	0.367
Race known to be neither white nor black, or unknown	Ref. cell	
Indicator whether participant and applicant for benefits are in same zip code (PDZIPSAME)		
Applicant and participant live in same zip code	0.607	0.317
Applicant and participant live in different zip code	0.965*	0.459
No information	Ref. cell	
Identity of payee relative to participant (REPREPAYEE)		
A family member received benefits on behalf of participant	0.929*	0.391
Participant received benefit payments himself/herself, an institution received payments on behalf of participant, or information unknown	Ref. cell	
Participant recipient benefit type (SSI_SSDI)		
Both SSI and SSDI	0.588	0.310
SSI only	Ref. cell	
Disability diagnosis classification (DIG)		
Participant had mental disability	0.433	0.231
Participant had physical disability, or information unknown	Ref. cell	
Living situation of participant (LIVING)		
Participant lives alone	-0.839**	0.302
Participant does not live alone, or information unknown	Ref. cell	
Count of addresses on file (MOVE)		
Two addresses on file	-0.768*	0.297
Three or more addresses on file	-1.760**	0.379
One address on file or no information	Ref. cell	
Count of phone numbers on file (PHONE)		
Three or more phone numbers on file	-1.171**	0.291
No information	Ref. cell	
Retirement destination county (CNTYRETIRE)		
Retirement destination county	-0.485	0.315
County that doesn't have this attribute	Ref. cell	

[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% and 1% levels respectively.

Table D.8. Variables in the Cooperation Logistic Propensity Model Ticket Participant Sample, Traditional Payment System

Main Effects	Parameter Estimate[a]	Standard Error
Variables in the Cooperation Model for Participants Using the Traditional Payment System		
Geographic region (based on U.S. Census divisions) of beneficiary's place of residence (DIVISION)		
East South Central	-1.159**	0.324
Middle Atlantic	0.358	0.314
Middle Atlantic	Ref. cell	
Participant's type of claim (TOC)		
Disability claim	0.192†	0.192
Survivor claim, or unknown	Ref. cell	
Disability diagnosis classification (DIG)		
Participant had physical disability (excluding deaf cases)	-0.639*	0.292
Participant had a mental disability, was deaf, or information about disability not given	Ref. cell	
Race of the participant (RACE)		
Black	0.569†	0.311
Race known to be neither white nor black, or unknown	Ref. cell	
Count of addresses on file (MOVE)		
One address on file	0.579	0.360
Two addresses on file	-0.170	0.401
Three or more addresses on file	-0.955*	0.446
No information	Ref. cell	
Count of phone numbers on file (PHONE)		
One or two phone numbers on file	0.592**	0.214
Three or more phone numbers on file	0.516*	0.212
No information	Ref. cell	
Racial/ethnic makeup of county (CNTYRACE)		
County with at least 90% non-Hispanic white population	0.325	0.182
County that doesn't have this attribute	Ref. cell	
County with population loss (CNTYPOPLOSS)		
County with population loss	-0.392	0.344
County that doesn't have this attribute	Ref. cell	
Low education county (CNTYLOWEDUC)		
County with low levels of education	-0.420*	0.185
County that doesn't have this attribute	Ref. cell	
Two- factor interactions[b]		
TOC*RACE		
Disability claim * Black	0.569	0.395

[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% and 1% levels respectively.

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

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

APPENDIX E

SUDAAN PARAMETERS FOR NATIONAL ESTIMATES FROM THE TTW ROUND 4 SAMPLE

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

```
proc descript data="SASdatasetname" filetype=sas design=wr;
nest       A_STRATA A_PSU / missunit;
weight    "weight variable" ;
subpopn     "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";
```

WEIGHT VARIABLES USED FOR CROSS- SECTIONAL ESTIMATES

Beneficiary sample: **Wtr4_ben**
Participant sample: **Wtr4_par**
Combined samples: **Wtr4_com**

NEST VARIABLES USED FOR CROSS- SECTIONAL ESTIMATES

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
   a. A_STRATA = 1410 nonSVRA EN participants in PSUs, Phase 1 states
   b. A_STRATA = 1420 nonSVRA EN participants not in PSUs, Phase 1 states
   c. A_STRATA = 1510 SVRA Acting as EN participants in PSUs, Phase 1 states
   d. A_STRATA = 1520 SVRA Acting as EN participants not in PSUs, Phase 1 states
   e. A_STRATA = 2410 nonSVRA EN participants in PSUs, Phase 2 states
   f. A_STRATA = 2420 nonSVRA EN participants not in PSUs, Phase 2 states
   g. A_STRATA = 2510 SVRA Acting as EN participants in PSUs, Phase 2 states
   h. A_STRATA = 2520 SVRA Acting as EN participants not in PSUs, Phase 2 states
   i. A_STRATA = 3410 nonSVRA EN participants in PSUs, Phase 3 states
   j. A_STRATA = 3420 nonSVRA EN participants not in PSUs, Phase 3 states
   k. A_STRATA = 3510 SVRA Acting as EN participants in PSUs, Phase 3 states
   l. A_STRATA = 3520 SVRA Acting as EN participants not in PSUs, Phase 3 states

A_PSU

1. Clustered sample cases both beneficiaries and participants cross-sectional samples

   A_PSU = PSU identifier

2. Unclustered sample cases in Milestone-outcome and Outcome-only Phase 2 participants and Outcome-only Phase 3 participants cross-sectional samples

   A_PSU = MPR_ID

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.

MATHEMATICA
Policy Research

www.mathematica-mpr.com

Improving public well-being by conducting high-quality, objective research and surveys

Princeton, NJ ■ Ann Arbor, MI ■ Cambridge, MA ■ Chicago, IL ■ Oakland, CA ■ Washington, DC

Mathematica® is a registered trademark of Mathematica Policy Research