National Beneficiary Survey Round 3 (Volume 3 of 3): User's Guide for Restricted and Public Use Files

Final Report

January 15, 2010

Debra Wright
Eric Grau
Yuhong Zhang
Danna Basson
Frank Potter
Kirsten Barrett

MATHEMATICA
Policy Research, Inc.

Contract Number:
0060-03-60129

Mathematica Reference Number:
8978

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

Submitted by:
Mathematica Policy Research
600 Maryland Avenue, SW
Suite 550
Washington, DC 20024-2512
Telephone: (202) 484-9220
Facsimile: (202) 863-1763
Project Director: Debra Wright

National Beneficiary Survey Round 3 (Volume 3 of 3): User's Guide for Restricted and Public Use Files

Final Report

January 15, 2010

Debra Wright
Eric Grau
Yuhong Zhang
Danna Basson
Frank Potter
Kirsten Barrett

MATHEMATICA
Policy Research, Inc.

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.

CONTENTS

Chapter **Page**

Chapter **Page**

Chapter **Page**

APPENDICES

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

TABLES

TABLES *(continued)*

TABLES *(continued)*

TABLES *(continued)*

ACRONYMS

ADLs:	Activities of Daily Living
AIC:	Akaike's Information Criterion
CAPI:	Computer-assisted personal interviewing
CATI:	Computer-assisted telephone interviewing
CDR:	Continuing Disability Review
CHAID:	Chi-Squared Automatic Interaction Detector
ENs:	Employment Networks
IADLs:	Instrumental Activities of Daily Living
ICD-9:	International Classification of Diseases – 9th revision
ICPSR:	Inter-University Consortium for Political and Social Research
IWP:	Individual Work Plan
MIE:	Medical Improvement Expected
MPR:	Mathematica Policy Research
MSA:	Metropolitan Statistical Area
NAICS:	North American Industry Classification System
NBS:	National Beneficiary Survey
PMSA:	Primary Metropolitan Statistical Area
PSU:	Primary Sampling Units
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 Units

STATA: Statistical software (STATA is a registered trademark of StataCorp LP, College Station, TX.)

SVRA: State Vocational Rehabilitation Agency (also called VRA or VR)

TTY: Teletypewriter

TTW: Ticket to Work

TRS: Telecommunications Relay Service

I. INTRODUCTION

As part of an evaluation of the Ticket to Work and Self-Sufficiency program (TTW), Mathematica Policy Research (MPR) conducted the third round of the National Beneficiary Survey (NBS) in 2006. The survey, sponsored by the Social Security Administration's (SSA) Office of Disability and Income Security Programs, collected data from a national sample of SSA disability beneficiaries (hereafter referred to as the Representative Beneficiary Sample) and a sample of TTW participants (hereafter referred to as the Ticket Participant Sample). The Ticket Participant Sample contains cross-sectional and longitudinal components, both of which are discussed in this report. MPR collected data using computer-assisted telephone interviewing (CATI) with computer-assisted personal interviewing follow-ups of CATI nonrespondents and those who preferred or needed an in-person interview to accommodate their disabilities.

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

The TTW program was implemented in three phases. In Phase 1, which began in February 2002, the program was rolled out in 13 states across the country. In Phase 2, which began in November 2002, the program was extended to an additional 20 states plus the District of

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

Columbia. Phase 3, which began in November 2003, implemented TTW in the remaining 17 states and U.S. territories (Thornton, et al. 2004).

A. OVERVIEW OF THE NATIONAL BENEFICIARY SURVEY

1. Survey Objectives

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

The NBS has five key objectives:

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

Round 3 NBS data will be combined with SSA administrative data to provide critical information on access to jobs and on employment outcomes for beneficiaries, including those who participate in the TTW program and those who do not. Though some sections of the NBS target beneficiary activity directly related to TTW, most of the survey captures more general information on SSA beneficiaries, including their disabilities, interest in work, use of services,

and employment. As a result, SSA and external researchers who are interested in disability and employment issues can use the survey data for other policymaking and program-planning efforts.

2. Data Collection Overview

Round 3 CATI data collection for both samples began in February 2006. Beginning in May 2006, MPR conducted in-person CAPI interviews with beneficiaries who did not respond to the CATI interview, as well as those who could not be located (and whose names and other information were sent to field interviewers for additional locating), or who requested an in-person interview to facilitate their participation in the survey. The survey instrument was identical in each mode. When possible, the interview was attempted with the sample person. If the sample person was unable to complete either a telephone or in-person interview, a proxy respondent was sought. Proxy interviews were attempted only when the sample member was unable to complete the survey himself or herself due to his/her disability. To promote response among Hispanic populations, the questionnaire was available in Spanish. For languages other than English and Spanish, interpreters conducted the interviews. A number of additional accommodations were made available for those with hearing and/or speech impairments including teletypewriter (TTY), Telecommunications Relay Service (TRS), amplifiers, and instant messaging technology.

As shown in Table I.1, the NBS round 3 sample comprised 3,382 cases selected for the Representative Beneficiary Sample and 5,697 cases selected for the Ticket Participant Sample. The Ticket Participant Sample includes 1,466 persons who were selected in round 1 (The Phase 1 Ticket Participant Longitudinal Sample), 1,350 persons selected in round 2 (The Phase 2 Ticket Participant Longitudinal Sample), and 2,881 Ticket participants who were newly selected for round 3 (1,508 from the Phase 2 cohort of participants and 1,373 from the Phase 3 cohort) . Ninety-five percent of the 1,350 Ticket Participants sampled at round 2 (1,289) were still signed

up with TTW at the time of the round 3 sample selection and were therefore considered part of the Phase 2 Cross-Sectional Sample. The remainder were not TTW participants at round 3 and were therefore not eligible for the Phase 2 Ticket Participant Cross-Sectional Sample (although they were still part of the Phase 2 Ticket Participant Longitudinal Sample). Phase 1 sample members were not part of the round 3 Ticket Participant Cross-Sectional Sample.

TABLE I.1

SAMPLE SIZES, TARGET COMPLETES, AND ACTUAL COMPLETES FOR CROSS-SECTIONAL SAMPLE

Sampling Strata	Sample Size	Target Completes	Actual Completes
Representative Beneficiary Sample	3,382	2,400	2,508
Ticket Participant Sample	4,170	3,000	3,115
Phase 2 Cohort	2,797	2,000	2,062
Phase 3 Cohort	1,373	1,000	1,053
Total Sample Size	**7,552**	**5,400**	**5,623**

Source: NBS, round 3.

The round 3 CATI and CAPI data collection was completed in September 2006. In the cross-sectional samples, 5,623 cases were completed (including 29 partially completed interviews)—2,508 from the Representative Beneficiary Sample and 3,115 from the Phase 2 and Phase 3 Ticket Participant cohorts. [2] An additional 982 cases were not eligible for the Ticket Participant Cross-Sectional Sample, but are included in the data file as completed interviews from the Phase 1 and Phase 2 Longitudinal Ticket Participant Samples (for a total of 4,097

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

Ticket Participant complete interviews)[3]. Thus the total number of completes is 6,605: 2,508 from the Beneficiary Sample and 4,097 TTW Participant completes (3,115+982). An additional 215 beneficiaries and 46 TTW participants were determined to be ineligible for the survey.[4] Across both samples, 5,104 cases were completed by telephone, and 1,501 were completed by CAPI. Proxy interviews were completed for 1,286 sample members (19 percent). There were 146 cases in which the sample member was unable to participate and a proxy could not be identified. The weighted response rate for the Representative Beneficiary Sample was 81.1 percent. The weighted response rate for the Ticket Participant Cross-Sectional Sample was 84.4 percent. The weighted response rate for the longitudinal Phase 1 TTW participants completing all three rounds was 62.8 percent. More information about the sample selection and sampling weights can be found in Chapter VI.

B. NBS RESTRICTED USE AND PUBLIC USE FILES

This guide describes the content and format of the NBS Restricted and Public Use data files and codebooks. To protect the anonymity of the respondents while providing accurate and detailed data, the NBS data are presented in two formats: a Restricted Use file available only to users approved by SSA and for use on specific research projects, and a Public Use File, planned to be released by SSA, for the general public to use in various statistical analyses. These two files present the same survey results, but offer differing degrees of accessibility to confidential information. For both data files, any information that could directly or indirectly identify a

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

[4] Ineligible sample members include those who were deceased, incarcerated; or no longer living in the continental United States; or those whose benefit status was pending. For the Ticket Participant Sample, ineligibles also included sample members who left the program after sampling was completed (although those who were in the round 1 sample and subsequently left the program were eligible for the Phase 1 longitudinal sample and those who were in the round 2 sample and subsequently left the program were eligible for the Phase 2 longitudinal sample).

respondent has been removed; this information includes respondents‘ names, Social Security numbers, and addresses. Because of its more widespread availability, the Public Use File has undergone extensive masking and has fewer available variables than the Restricted Use File. Even with the variables masked, however, the NBS Public Use File offers a wide variety of pertinent variables and topics for the general public to use. The masking procedures employed to create the Public Use File are discussed in more detail in Chapter V. A listing of the variables available on the NBS Restricted Use Data File and Public Use Data File is included as Appendix A.

The Public Use File will be available to researchers through the Health and Medical Care Archive at the Inter-University Consortium for Political and Social Research (ICPSR). The file can be downloaded directly from the ICPS Web site (www.icpsr.umich.edu). Researchers must contact SSA to obtain permission to use the Restricted Use File.

C. ROUND 3 DATA DOCUMENTATION REPORTS

The following reports make up the complete documentation describing the NBS, the round 3 data collection, and the data files:

- ***Editing, Coding, Imputation, and Weighting Report (Grau, et al. 2008).*** This report summarizes the editing, coding, imputation, and weighting procedures as well as the development of standard errors for the round 3 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 the Ticket Participant Sample (and describes the procedures for combining these samples); describes the procedures used to impute missing responses; and discusses procedures that should be used to estimate sampling variances for the NBS.
- ***Cleaning and Identification of Data Problems Report (Wright and Barrett 2008).*** This report describes the data processing procedures performed for round 3 of the NBS. It outlines the data coding and cleaning procedures and describes the data problems identified, their origins, and the corrections implemented to create the final data file. The report describes the data issues by sections of the interview and

concludes with a summary of types of problems encountered and general recommendations.

- ***User's Guide for Restricted and Public Use Data Files (current report).*** This report is designed to provide users with information about the restricted and public use data files including construction of the files; weight specification and variance estimation; masking procedures employed in the creation of the Public Use File; and a detailed overview of the questionnaire design, sampling, and NBS data collection. The report also contains some information covered in the two reports mentioned above, including procedures for data editing, coding of open-ended responses, and variable construction; and a description of the imputation and weighting procedures and development of standard errors for the survey.

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

- ***NBS Questionnaire.*** This document contains all items on the round 3 survey and includes documentation of skip patterns, question universe specifications, text fills, interviewer directives, and consistency and range checks.
- ***NBS Restricted Access and Public Use File Codebooks.*** The codebooks provide extensive documentation for each variable on the file including variable name, label, position, variable type and format, question universe, question text, number of cases eligible to receive each item, constructed variable specifications, and user notes. Frequency distributions and means are also included, as appropriate.

In the discussion that follows, we provide detailed information about the NBS to assist users of the round 3 public and restricted access data files. Chapter I offers an overview of the NBS and study objectives. In Chapter II we describe the sample design of the NBS. Chapter III provides an overview of the questionnaire design. Chapter IV explains the NBS data collection, including the locating and calling protocols for this survey. Chapter V is devoted to variable construction and editing, and provides information on the coding of verbatim and open-ended responses found in the NBS. Masking procedures employed in the creation of the Public Use File are also discussed. Chapter VI explains the weighting, imputation, and variance estimates used in this survey. Finally, Chapter VII provides information about using the NBS data files, including weight specification and variance estimation.

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

II. SAMPLE DESIGN

A. OVERVIEW OF THE DESIGN

SSA implemented the TTW program in three phases spanning three years, with each phase corresponding to about one-third of the states. The initial NBS survey design called for four national cross-sectional surveys (called rounds) of Ticket-eligible SSA disability beneficiaries—one each in 2003, 2004, 2005, and 2006—and cross-sectional surveys of Ticket participants in each of three groups of states (Phase 1, Phase 2, and Phase 3 states)—defined by the year in which the program was rolled out (Bethel and Stapleton 2002).[5] This design was subsequently revised to accommodate Phase 1 data collection starting in 2004 rather than 2003. In addition, the final round was postponed to address the experiences of TTW participants under the new TTW regulations, implemented in July 2008. The fourth round will include a cross-sectional Representative Beneficiary survey as well as a survey of new Ticket Participants and is planned for 2009. Details of the sample design for round 4 are not yet determined; in a change from the original design, Ticket participants from previous rounds will not be re-interviewed.

One group of sample members in each of the first two cross-sectional surveys of Ticket participants was followed longitudinally across rounds: Phase 1 sample members who were active in the TTW program in round 1, and Phase 2 sample members active in round 2. The

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

original sample design called for re-interviewing only those longitudinal cases that had completed the previous round. However, based on MPR's recommendation, re-interviews were attempted with all longitudinal cases. Table II.1 gives the original planned sample sizes for all rounds of data collection. The initial sampling and survey design documents are available from SSA upon request.

TABLE II.1

NATIONAL BENEFICIARY AND TTW PARTICIPANT SAMPLE SIZES

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

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

[a] Sample sizes refer to number of completed interviews

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

[c] The All Years 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 upon the number of completed interviews for the sample member in question.

In round 1 (2004), two surveys were fielded: the first national survey of all beneficiaries (the Representative Beneficiary Sample) and the first cross-sectional survey of Ticket participants in the Phase 1 states (the Ticket Participant Sample). Three surveys were fielded in round 2 (2005):

1. The second national survey of all beneficiaries (The Representative Beneficiary Sample).

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

Additionally, we attempted to re-interview Phase 1 Ticket Participants who were selected into the sample at round 1, whether or not they had been interviewed in round 1 (the Phase 1 Longitudinal Sample).

Three surveys were also fielded in round 3 (2006):

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

In addition, we attempted to re-interview Phase 1 Ticket participants who were selected into the sample at round 1 whether or not they had been interviewed in rounds 1 or 2 (the Phase 1 Longitudinal Sample), and Phase 2 participants who were selected into the sample at round 2, whether or not they had been interviewed at round 2 (the Phase 2 Longitudinal Sample). Most, but not all of the Phase 2 longitudinal sample cases were also part of the Phase 2 Cross-Sectional Ticket Participant Sample.

In each first follow-up year (round 2 for Phase 1 participants, and round 3 for Phase 2 participants), a supplemental sample of those who had entered the TTW program since the first year of rollout for each phase, or otherwise had not been sampled before, was selected to produce an expanded second-year cross-sectional sample survey. For Phase 1 and Phase 2 participants,

this resulted in cross-sectional samples for two consecutive years. The cross-sectional surveys consisted of the supplemental cases, plus the longitudinal cases who were still Ticket participants at the time of sampling. In round 3, there were cross-sectional samples for Phase 2 and Phase 3 participants; however, no supplemental cross-sectional sample was selected for the Phase 1 Ticket participants at round 3. (Hence, no Phase 1 cross-sectional weights were calculated in round 3.)

The NBS used a multi-stage sampling design (which was used for all survey rounds) with a supplemental single-stage sample for some Ticket participant populations. For the multi-stage design, data from SSA on the counts of eligible beneficiaries in each county were used to form the primary sampling units (PSUs) consisting of one or more counties. A stratified national sample of 80 PSUs was selected; Los Angeles County and Cook County (Chicago) were selected with certainty because of the number of SSA beneficiaries in these counties. Because of the size of these two counties (in both beneficiary population and geographic area), Secondary Sampling Units (SSUs) were formed using zip codes of beneficiaries. Four SSUs were selected from Los Angeles County and two were selected from Cook County (Chicago). PSUs were selected with probability proportional to size of the beneficiary population in them. One PSU was selected twice because of the large number of beneficiaries in the included county, therefore the final number of PSUs selected was 79.[6] The sample of all SSA beneficiaries (the Representative Beneficiary Sample) was selected from among beneficiaries residing in these PSUs/SSUs using age-defined sampling strata. The final sample size for the Representative Beneficiary Sample in round 3 was 3,382 (see Table II.2 for a detailed description of sample size by stratum).

[6] For the data analysis, the number of PSUs was 80, the original number of selections.

TABLE II.2

ROUND 3 CROSS-SECTIONAL SAMPLE SIZES AND TARGET COMPLETES PER SAMPLING STRATUM

Sampling Strata	Sample Size	Target Completes	Actual Completes
Beneficiary Sample	3,382	2,400	2,508
18 to 29 Years Old	943	666	698
30 to 39 Years Old	941	666	672
40 to 49 Years Old	935	666	711
50 to 64 Years Old	563	402	427
Ticket Participant Cross-Sectional Sample	4,170	3,000	3,115
Phase 2 Cohort[a]	2,979	2,000	2,062[b]
Traditional Payment Type	867	666	733
Milestone-Outcome Payment Type (Unclustered)	548	666	306
Milestone-Outcome Payment Type (Clustered)	389		357
Outcome-Only Payment Type (Unclustered)	870	666	579
Outcome-Only Payment Type (Clustered)	123		87
Phase 3 Cohort	1,373	1,000	1,053
Traditional Payment Type	444	333	369
Milestone-Outcome Payment Type	444	333	362
Outcome-Only Payment Type (Unclustered)	248	333	144
Outcome-Only Payment Type (Clustered)	237		144
Total	**7,552**	**5,400**	**5,623**

Source: NBS round 3.

[a] The Phase 2 Cross-Sectional Sample included 1,289 Ticket participants sampled at round 2 who were still in SSA's Ticket Participant File at round 3 and 1,508 Ticket participants in SSA's Ticket Participant File who were newly sampled at round 3. Phase 1 sample members are not part of the round 3 Cross-Sectional Sample.

[b] There are 982 additional completed interviews among Ticket Participants on the round 3 data file (for a total of 4,097 completed interviews). These are Phase 1 and Phase 2 longitudinal sample members that were not eligible for the round 3 Cross-Sectional Sample (that is they were not TTW Participants at round 3).

These PSUs were also used to generate the Phase 2 and Phase 3 Ticket Participant Samples for each of the three Ticket Program EN payment types (outcome-only, milestone-outcome, and traditional vocational rehabilitation). Each phase of the Ticket program included only one-third of all states; therefore, the sample for each phase was based on only one-third of the PSUs. For participants in Phase 2 states using either the milestone-outcome or the outcome-only payment system and for participants in Phase 1 or Phase 3 states using the outcome-only payment system, the number of Ticket participants in the clusters was insufficient to support the analytic objectives of the survey, so the clustered samples were supplemented by an independent

unclustered sample of participants.[7] The clustered Ticket Participant Samples were selected using the same PSUs, but due to the small number of Ticket participants, the Secondary Sampling Units were not used and the sample was drawn from all participants in the PSUs.[8] For participants using the milestone-outcome and outcome-only payment types in Phase 2 states, and for participants using the outcome-only payment system in Phase 1 and Phase 3 states, the unclustered sample was a stratified random sample using two strata: participants in the PSUs and participants outside of the PSUs. This stratification was needed to control the sample release.

At round 3, the final sample size for the Ticket Participant Cross-Sectional Sample was 4,170. This consisted of the Phase 2 Ticket Participant Cross-Sectional Sample, with 2,797 sample members, and the Phase 3 Ticket Participant Cross-Sectional Sample with a sample size of 1,373 (see Table II.2).

The round 3 Phase 2 Ticket Participant Cross-Sectional Sample included 1,289 longitudinal cases, along with 1,508 Supplemental Sample cases that were interviewed for the first time in round 3. Of the 1,350 cases in the Phase I Longitudinal Sample, 1,289 (95 percent) were found in SSA's file of active Ticket participants at round 3. The remaining 61 were either deceased or were not TTW participants at round 3 and were therefore not eligible for the Phase 3 Ticket Participant Cross-Sectional Sample (see Table II.3 for longitudinal sample sizes and completes per stratum). The full sample of TTW participants included 5,697 sample members (4,170 participants in the cross-sectional samples, 1,466 Phase 1 sample members who were in the

[7] The use of two independent samples, called a paired sample design, was implemented for the Phase 2 and Phase 3 Ticket Participant Cross-Sectional Sample, and for the Phase 1 and Phase 2 Ticket Participant Longitudinal Sample.

[8] Participants of the Ticket program are also SSA beneficiaries and these samples of participants are designed to support the more detailed analysis required for the evaluation of the program. We anticipated that some Ticket participants would be selected in the beneficiary survey (and a small number of Ticket participants were selected in both samples).

Phase 1 longitudinal sample only, plus 61 sample members who were participants in round 2 and not at round 3, as noted above). The 1,466 sample members in the Phase 1 Ticket Participant Longitudinal Sample may or may not have been Ticket participants at round 3, but since Ticket participants from Phase 1 states who entered the TTW program after round 1 sampling were not sampled at round 3, they were not part of the round 3 Ticket Participant Cross-Sectional Sample.

TABLE II.3

ROUND 3 PHASE 1 AND PHASE 2 LONGITUDINAL SAMPLE SIZES AND TARGET COMPLETES PER SAMPLING STRATA

Sampling Strata	Sample Size	Target Completes	Actual Completes[a]
Phase 1, complete in rounds 1 and 2	1,466	922	897
Traditional Payment Type	441	307	304
Milestone-Outcome Payment Type	455	307	282
Outcome-Only Payment Type (Unclustered)	447	307	243
Outcome-Only Payment Type (Clustered)	123		68
Phase 1 complete in rounds 1, 2, and 3	1,466	850	767
Traditional Payment Type	441	283	266
Milestone-Outcome Payment Type	455	283	241
Outcome-Only Payment Type (Unclustered)	447	283	196
Outcome-Only Payment Type (Clustered)	123		64
Phase 2, complete in rounds 2 and 3	1,350	922	831
Traditional Payment Type	437	307	308
Milestone-Outcome Payment Type (Unclustered)	220	307	119
Milestone-Outcome Payment Type (Clustered)	216		146
Outcome-Only Payment Type (Unclustered)	391	307	210
Outcome-Only Payment Type (Clustered)	86		48
Total Sample Size	**2,816**		

Source: NBS round 3.

[a] No totals are possible for target and actual completes since the definition of a completed case is dependent on how many rounds of data are included in the longitudinal analysis. Target completes for each payment type are based on a 7.8 percent attrition rate per round.

B. TARGET POPULATION

The target population for both the Representative Beneficiary Sample and the Ticket Participant Sample consisted of SSI and SSDI beneficiaries between the ages of 18 and 64. For the Representative Beneficiary Sample, the target population included beneficiaries in all 50

states and the District of Columbia who were in active pay status as of June 2005. [9] There were two subpopulations of these beneficiaries who are not eligible for Ticket participation but were included in the survey samples to give complete coverage of the national beneficiary population:

- Beneficiaries who were designated as Medical Improvement Expected (MIE) at the time they received their allowances and who had not yet completed a first Continuing Disability Review (CDR)
- Young SSI recipients who were receiving benefits because of their eligibility as a child, and were in the process of completing a re-determination under the adult eligibility criteria.

The beneficiary target population included approximately 10.4 million persons; approximately 2.1 million beneficiaries were in the sampled PSUs.

For the Ticket Participant Cross-Sectional Sample, the target population included beneficiaries who had used the Ticket at least once as of January 1, 2005, or between January 1, 2005, and October 2, 2005. For the Ticket participants, the study population was constrained by the TTW rollout schedule. For cross-sectional estimates, the target population for the round 3 survey included beneficiaries who were participants in SSA's TTW program in the Phase 2 or Phase 3 rollout states. Participants were assigned to a phase for this study on the basis of their address at the time of program rollout regardless of their current address. Thus, a Phase 2 participant (middle rollout state) might reside in any state at the time of the survey. At the time of round 3 sampling, the target population for the Phase 2 Ticket Participant Cross-Sectional Sample included 33,500 Ticket participants and the target population for the Phase 3 Ticket Participant Cross-Sectional Sample included 31,023 Ticket participants.

[9] Beneficiaries in the Trust Territories and Puerto Rico were excluded from the survey target population.

For longitudinal estimates, the target population is defined by the population of Ticket participants at the time of round 1 sampling for Phase 1 sample members, and the population of Ticket participants at the time of round 2 sampling for Phase 2 sample members. The target population for the Phase 1 longitudinal sample included 21,477 Ticket participants (as of round 1) and the target population for the Phase 2 longitudinal sample included 21,196 Ticket participants (as of round 2).

For the Phase 1 Ticket participants, the samples were designed for the analysis of one longitudinal population: persons who were participants at the time of the first data collection. For the Phase 2 Ticket participants, the samples were designed for the analysis of two overlapping populations:

- The longitudinal population: persons who were participants at the time of the round 2 data collection; and
- The cross-sectional population: persons who were currently participants

Phase 3 Ticket participants were selected for the first time in round 3 from the Ticket participant sampling frame.

C. PRIMARY SAMPLING UNIT FORMATION AND SELECTION

PSUs were needed for both the Representative Beneficiary Survey and the Ticket Participant Survey and were constructed using county-level beneficiary counts. Based on the design report for the TTW Evaluation (Bethel and Stapleton 2002), the design for the Representative Beneficiary Survey called for 80 PSUs to be formed from counties or groups of counties. The design report also recommended that in the geographically largest PSUs, Secondary Sampling Units (SSUs) would be formed based on zip codes, and a sample of these would be selected. The clustered Ticket Participant Sample was selected in the same manner as the Representative Beneficiary Sample using the same PSUs, but due to the small number of Ticket participants, the

Secondary Sampling Units were not used and the sample was drawn from all participants in the PSUs. To construct the PSUs, county-level counts of beneficiaries in four age strata (18-29, 30-39, 40-49, and 50-64) and a composite size measure were used (Folsom et al. 1987). The composite size measure incorporates the count of beneficiaries and the desired sampling rate of beneficiaries in each age stratum and permits equal probability of selection of beneficiaries within each age stratum across PSUs, and approximately equal workload in each PSU. To form the PSUs, counties were ordered within each state by geography using a score based on latitude and longitude. An eligible PSU needed a composite size measure above a specific level to ensure that adequate counts of beneficiaries existed in each of four sampling strata. The PSUs were also evaluated on the basis of geographic size (square miles), topography (lakes, rivers, and mountain ranges) and transportation access among counties in a PSU (roadways in mountainous areas and bridges around the Great Lakes).

In total, 1,330 PSUs were formed with 48 percent (664 PSUs) having a single county; 84 percent (1,118 PSUs) had three or fewer counties. Of the 1,330 PSUs, just 30 (2.2 percent) included 10 or more counties; mostly rural areas in Western states.

For sample selection of PSUs, the PSUs were stratified explicitly by the Ticket program‘s three phases (each accounting for approximately one-third of the states). Based on the selection frequencies for the PSUs computed using the composite size measure, two PSUs were classified as certainty PSUs selections (Los Angeles County and Cook County (Chicago)). Because of the size of the Los Angeles County selection frequency, this PSU was allocated twice the sample size allocated to the others. To complete the sample of 80 PSUs, we selected 77 other noncertainty PSUs with probability proportional to the composite size measure within each Ticket phase stratum. The selection of the PSUs was controlled implicitly by SSA region, state within SSA region, and a beneficiary weighted score (from 0 to 9) based on the 2003 Urban

Influence Code (Area Resource File 2003). In the Phase 1 states, 23 PSUs were selected and in the Phase 2 states, 25 PSUs were selected. In the Phase 3 states, 31 PSUs were selected. (As noted previously, the Los Angeles County PSU accounted for two PSU selections).

In the Los Angeles and Chicago certainty PSUs, Secondary Sampling Units were formed using counts of beneficiaries in each stratum for five-digit zip codes and the composite size measure. Once again, SSUs consisted of one or more zip code areas such that the aggregate composite size measure exceeded the criterion value. In the Los Angeles PSU, 62 SSUs were formed and four were selected with probability proportional to the composite size measure. In the Chicago PSU, 44 SSUs were formed and two were selected with probability proportional to the composite size measure.

D. STRATA DEFINITIONS AND SAMPLE SIZES

The sample is designed to be statistically and operationally efficient and to provide adequate sample sizes for the planned analyses. In order to ensure a sufficient number of persons seeking work, the Representative Beneficiary Sample was classified into sampling strata based on age, with persons in the younger age categories selected at higher rates than those in the oldest age category. The Representative Beneficiary Sample was divided into the following age groups: 18-29, 30-39, 40-49, and 50-64, which were used as the sampling strata. The target number of completed interviews for round 2 was 1,333 beneficiaries in each of the three younger age groups (18-29, 30-39, and 40-49). For the 50-64 age cohort, the target number of completed interviews was 800 beneficiaries.

The sampling strata for the Ticket Participant Samples were defined by the payment system. For Ticket participants, services received from ENs can be provided under three program payment systems: (1) outcome-only, (2) milestone-outcome, or (3) the traditional vocational rehabilitation reimbursement system. (See the –Evaluation of the Ticket to Work Program Initial

Report," Thornton, et al. (2004), for more information about the EN program payment systems.) Because the use of the outcome-only and milestone-outcome payment systems was low among Ticket participants, both a clustered and unclustered sample of participants was selected for each of these payment types. The sample of participants using the traditional payment type was limited to a clustered sample. The target number of completed interviews for participants in the cross-sectional samples at round 3 was 3,000 overall, with a target of approximately 2,000 for the Phase 2 cohort (666 in each payment type) and 1,000 for the Phase 3 cohort (333 in each payment type).

In order to statistically combine the clustered and unclustered samples, we needed to establish comparability between the portions of the samples related to the data collection effort because, while both samples received central office locating and telephone interviewing, only the clustered sample received field locating and in-person interviewing. Sample members in both the clustered and unclustered samples underwent the same level of central office locating activities (including batch processing through search databases and individualized locating efforts) to identify a telephone number so that a telephone interview could be attempted. For the unclustered sample, participants who could not be located or who required an in-person interview were considered nonrespondents. Operationally, these cases were "closed out" and classified as ineligible for purposes of sampling weight computation.[10] For the clustered sample, beneficiaries who could not be located or who required an in-person interview were eligible for a field follow-up and were assigned to field locators/interviewers. The sample members in both the clustered and unclustered samples were comparable up to the point of assignment of sample members for field work. The samples from the clustered and unclustered sample before

[10] They were treated differently, however, than other ineligible cases, which were operationally treated as respondents for the purposes of calculating sample weights and response rates.

assignment for field work could be statistically combined because the two samples represented the same subpopulation (Ticket participants who could be located by central office locating efforts and interviewed by telephone). The sample members in the clustered sample who were assigned for field work represented a subsample of sample members representing the subpopulation who required field work for locating and interviewing.

For fielding purposes in both the Representative Beneficiary Sample and the Ticket Participant Samples, we selected many more cases than we needed (called the augmented sample) to ensure that an adequate pool of sample would be available if we found that the response and eligibility rates during data collection differed from our initial assumptions. Within each stratum, an equal probability sample of beneficiaries or participants was selected using a sequential selection algorithm with the sampling frame sorted by disability diagnosis, race/ethnicity, and zip code. These sorting factors ensured an approximate proportional allocation of the sample across levels of these factors and, therefore, enhance the face validity of the sample across these factors.

For the Representative Beneficiary Sample, we selected for the augmented sample approximately 4,000 beneficiaries in each of the three younger age groups (18-29, 30-39, and 40-49) and 3,000 beneficiaries in the oldest age cohort. For Phase 1 Ticket participants, all sample members that had been sampled in round 1 were included in the sample for round 3. Similarly, for Phase 2 Ticket participants, all sample members that had been sampled in round 2 were included in the sample for round 3. However, Phase 1 sample members were not included in the round 3 Ticket Participant Cross-Sectional Sample, since Ticket participants in Phase 1 sates who joined the Ticket program after round 1 were not sampled at round 3. For the round 3 cross-sectional sample, we selected a supplemental sample of Phase 2 TTW participants from two sources: (1) Phase 2 Ticket participants in round 2 who had not been selected for the round 2

sample and (2) Phase 2 Ticket participants who started participation after the round 2 sampling file was developed. The size of the augmented supplemental sample was sufficiently large to ensure approximately 2,000 target completes in the Phase 2 sample, though it varied by payment type according to the number of additional Phase 2 sample members needed after accounting for longitudinal cases. For augmented samples of the Phase 3 Ticket participants, we selected approximately 666 participants in each payment-type stratum. These augmented samples were randomly partitioned into subsamples (called "waves") to allow controlled release of sample throughout the data collection effort. During the data collection period, we monitored the sample results and determined whether, and in which strata and PSUs, additional waves of sampled cases were needed.

III. QUESTIONNAIRE DESIGN

The NBS collects data on a wide range of topics including employment, disability, experience with a variety of SSA programs, employment services used in the past year, health and functional status, health insurance, income and other assistance, and sociodemographic information. The survey items were developed and initially pre-tested as part of a separate contract held by Westat. Revisions were made by MPR to prepare the instrument for CATI/CAPI programming, and additional minor wording changes were made after pretesting. The survey instrument is available from MPR upon request.

To promote response among Hispanic populations, the questionnaire was translated into Spanish. In some cases, because the Spanish speaker was more familiar with a word or term in English than in Spanish, the term was provided in both languages so that interviewers could reinforce the question by using the second language as a probe, if necessary.[11] Measurements were treated in a similar way. Thus, questions that mentioned a specific weight also mentioned the kilogram equivalent.[12] Interpreters were used as needed to conduct interviews in languages other than Spanish.

A. QUESTIONNAIRE SECTIONS

The questionnaire is divided into 13 sections, labeled A through M:

- Section A–Introduction and Screener

[11] For example, on item L-5: Did {you/NAME} receive any food stamps last month? Spanish: Recibió {usted/NAME} food stamps o cupones de alimentos el mes pasado?

[12] For example, on item Jb-10: {Do you/Does NAME} have any difficulty lifting and carrying something as heavy as 10 pounds, such as a full bag of groceries? Spanish: Tiene {usted/NAME} cualquier dificultad en levantar y cargar algo que pesa hasta unas 10 libras {4½ kilos}, tal como una bolsa llena con compras del mercado?

- Section B–Disability and Current Work Status
- Section C–Current Employment
- Section D–Jobs/Other Jobs During 2005
- Section E–Awareness of SSA Work Incentive Programs and Ticket to Work
- Section F–Ticket Non-Participants in 2005
- Section G–Employment-Related Services and Supports Used in 2005
- Section H–Ticket Participants in 2005
- Section I–Health and Functional Status
- Section J–Health Insurance
- Section K–Income and Other Assistance
- Section L–Sociodemographic Information
- Section M–Closing Information and Observations.

Detailed descriptions of each section are provided below:

1. Section A–Introduction and Screener

This section confirms that the correct sample person has been contacted and verifies that the sample person is still eligible for the survey. In addition, the screener allowed interviewers to:

- Identify any barriers to participation and, if needed, identify a proxy respondent. The sample member was offered every opportunity to complete the interview himself or herself, and a proxy was only accepted if necessary.
- Identify the need for an interpreter for a respondent who spoke a language other than English or Spanish.
- Administer a cognitive assessment to ensure that the respondent would be capable of completing the survey.

Due to the complexity of the survey, a cognitive assessment was administered to respondents (both sample persons and proxy respondents) prior to the interview. Respondents were read three questions (a brief description of what it meant that the survey was confidential,

what it meant that the survey was voluntary, and an overview of the study topics) and asked to reiterate the concepts in his or her own words. If the respondent was not able to restate a concept, the question was read a second time. If the respondent could not restate a concept after being asked a second time, he or she was asked if there was someone else who could answer questions about his or her health, daily activities, and any jobs he or she might have (such as a friend, parent, caseworker, or payee). An interview was then pursued with the proxy respondent. To minimize bias in reporting, attitudinal and opinion items were skipped: Proxy respondents were not asked to provide subjective assessments on behalf of the sample person; for example, regarding satisfaction with jobs or programs. The constructed variable C_Rtype indicates whether the sample person or a proxy completed most of the interview.

2. Section B–Disability and Current Work Status

This section collects information on the beneficiary‘s limiting physical or mental condition(s) and current employment status. If the beneficiary is not currently employed, the section explores reasons for not working. This section also includes questions designed to determine the job characteristics that are important to beneficiaries and collects information about work-related goals and expectations.

3. Section C–Current Employment

Questions in this section collect detailed information about the beneficiary‘s current job(s). Respondents are asked about the type of work performed, type of employer, hours worked, benefits offered, and wages earned. The section also asks about work-related accommodations received, as well as those needed but not received. Other questions solicit information about job satisfaction.

4. Section D–Jobs/Other Jobs During 2005

This section collects information about employment during the 2005 calendar year, including type(s) of employer(s), hours worked, wages earned, and reasons for leaving employment, if applicable. Other questions ask whether beneficiaries worked or earned less than they could have (and if so, the reasons why), and collect information about experiences related to Social Security benefit adjustments due to work.

5. Section E–Awareness of SSA Work Incentive Programs and Ticket to Work

This section includes questions designed to assess whether the beneficiary is aware of, or is participating in, specific SSA work incentive programs and services. For the TTW program, information is collected on how beneficiaries learned about the program, the names of the providers they signed up with, and dates they signed up with their service providers.

6. Section F–Ticket Non-Participants in 2005

This section is administered to beneficiaries not participating in the TTW program and collects data on reasons for nonparticipation. It asks whether the beneficiary has attempted to learn about employment opportunities (including TTW), problems the beneficiary may have had with Employment Networks or other employment agencies, and how those problems were handled or resolved.

7. Section G–Employment-Related Services and Supports Used in 2005

Questions in this section ask beneficiaries about their use of employment-related services in calendar year 2005, including the types of services received, the types of providers used, how long they received services, how the services were paid for, and reasons for and satisfaction with service utilization. Other questions ask about sources of information about services and the nature of any services that were needed but not received.

8. Section H–Ticket Participants in 2005

This section asks 2005 Ticket to Work participants about their experiences with the program, including information related to their decision to participate in the Ticket program, the kinds of information they used to pick their current service providers, development of the individual work plan (IWP), and any problems experienced with services provided by an Employment Network. The section also includes a series of questions about how problems with Employment Networks were resolved and overall satisfaction with the Ticket to Work program.

9. Section I–Health and Functional Status

This section includes questions about the beneficiary‘s health status and everyday functioning, including the need for special equipment or assistive devices. Information is solicited regarding general health status (via the SF-8™ scale), [13] difficulties with Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs), a variety of functional limitations, substance abuse/dependence (using the CAGE Alcohol Abuse Screener),[14] and treatment for mental health conditions.

10. Section J–Health Insurance

Questions in this section collect information about the sources of health insurance coverage, both at the time of interview and during calendar year 2005.

[13] SF-8™ is a trademark of QualityMetric, Inc.

[14] See Mayfield, D., McLeod, G., Hall, P. (1974). The CAGE questionnaire: Validation of a new alcoholism instrument. American Journal of Psychiatry 131, 1121-1123.

11. Section K–Income and Other Assistance

Questions in this section ask about sources of income, including income received from earnings, Social Security, workers‘ compensation, and other government programs and sources.

12. Section L–Sociodemographic Information

This section collects basic demographic information about the beneficiary, such as race, ethnicity, education, parental education, marital status, living arrangements, and household income.

13. Section M–Closing Information and Observations

In this section, address information is collected for the sample person, and telephone information for up to two contact people is collected for participants who may be selected for future survey rounds. The interviewer also records reasons a proxy or assistance was required, if appropriate, and documents special circumstances.

B. QUESTIONNAIRE PATHING AND RESPONDENT TYPE

Sample members in the Representative Beneficiary Sample and the Ticket Participant Sample received the same version of the NBS questionnaire. Pathing to questions about participation in the TTW Program was not based on sample type, but rather to answers given to items in previous sections (awareness of the program and use of the Ticket). Similarly, both CATI and CAPI respondents received the same questionnaire.

All respondents were asked questions from sections A, B, E, G, I, J, K, L, and M. Only respondents who reported that they were currently working were asked questions from section C. Similarly, only respondents who reported working in 2005 were asked questions in section D. Section F was asked of respondents who reported that they had never tried to get a Ticket from SSA, had never tried to use a Ticket to sign up with a provider, or were not signed up with a

provider in 2005. Only respondents who reported using their Ticket to sign up with a provider in 2005 were asked questions from section H. See Table III.1 for a summary description of the main questionnaire pathing.

The NBS instrument, which is programmed in Blaise, is complex and involves numerous integrated skips, within and across sections. Further complexities in questionnaire pathing are introduced by the utilization of preloaded SSA administrative data and allowances for proxy participation. Preloaded data about respondents‘ disability-benefits status (SSI, SSDI, or both), the phase of TTW program roll-out, age at which they first received SSI benefits, and TTW participant status, determine pathing for certain survey items. Other administrative variables are used as fills at particular items to provide respondents with local names of programs or to prompt recognition of program participation. See Table III.2 for a complete list and description of preloaded variables. Phase of TTW roll-out was not included as a preload at round 3 since the item that referenced phase at round 1 was not included at round 2 or 3.

Finally, since proxies are necessary when the sample member‘s disability precludes participation, the instrument was programmed to fill the proper pronoun or name in the question text after the interviewer indicated who the survey respondent would be (sample member or proxy). Additionally, attitudinal and opinion items were skipped for proxy respondents so as to minimize bias in reporting. (See Table III.3 for a complete list of items that were not asked of proxy respondents.) Proxy interviews were completed for 1,286 cases.

TABLE III.1

NBS INSTRUMENT SECTIONS

Section	Title of Section	Respondents Receiving the Section
A	Screener	All respondents
B	Disability/Current Work Status	All respondents
C	Current Employment	Respondents who answer (B24 = YES). Question B24: Are you currently working at a job or business for pay or profit?
D	Jobs/Other Jobs During 2005	Respondents who answer (B30 = YES). Question B30: Did you work at a job or business for pay or profit anytime in 2005?
E	Awareness of SSA Work Incentive Programs and Ticket to Work.	All respondents
F	Ticket Non-Participants in 2005	Respondents who answer (E35 = NO, DON'T KNOW, OR REFUSED). Question E35: Did you ever try to get a Ticket from Social Security or anywhere else? OR Respondents who answer (E36 = NO, DON'T KNOW, OR REFUSED). Question E36: Have you ever used your Ticket to sign up with an Employment Network? OR Respondents who answer (E37 = NO, DON'T KNOW, OR REFUSED). Question E37: Were you signed up with any Employment Network or a State Vocational Rehabilitation Agency at any time in 2005?
G	Employment-Related Services and Supports Used in 2005	All respondents
H	Ticket Participants in 2005	Respondents who answer (E37 = YES) Question E37: Were you signed up with any Employment Network or a State Vocational Rehabilitation Agency at any time in 2005? OR Respondents who answered (round 2 E41 or round 1 E45 = YES) Question E41 and Question E45: Are you currently (in 2006) signed up with an Employment Network?
I	Health and Functional Status	All respondents
J	Health Insurance	All respondents
K	Income and Other Assistance	All respondents
L	Sociodemographic Information	All respondents
M	Closing Information and Observations	All respondents

Source: NBS, round 3.

TABLE III.2

SURVEY PRELOADS

Variable	Definition	Purpose
Bstatus	SSA benefit type (SSI only, SSDI only, or SSI and SSDI) received by sample member.	Used to determine pathing for awareness of SSA work incentive items. Only respondents who received SSDI benefits were asked items E3-E13. Only respondents who received SSI were asked items E15-E18.
DOB	Sample member date of birth.	Reported date of birth (or age) was matched with administrative data to verify that the correct person was contacted in the screener portion of the survey.
ENsample	Name of the Employment Network (EN) to which the sample member's ticket was assigned at the time the TTW Participant Sample was drawn.	Used as a fill at E24 to prompt TTW participants who responded that they had never heard of the TTW program. This item reminds respondents that according to SSA, the sample person's ticket was assigned to this EN (as of the date the sample frame was drawn).
LocalPAA	Name of Local Protection and Advocacy Group in the sample member's state of residence (as reported at time of survey).	Used at items H52, H53, H54, and H55 to identify, by name, the Protection and Advocacy Group in the respondent's area.
SDate	Date sample frame drawn for TTW participants.	Used as fill at E24 to prompt TTW participants who responded that they had never heard of the TTW program. This item reminds respondents that according to SSA, the sample person's ticket was assigned to an EN (as of the date the sample frame was drawn).
SSIage	Age at which sample member first received SSI benefits.	Used to determine pathing at items E11 and E12. Only respondents who received SSI before the age of 22 (and who were also 25 or younger) received these items.
StateMed	State name for Medicaid. Based on state of residence reported at time of survey.	Used at item J2 to identify, by name, the Medicaid program in the respondent's state.
Tstatus	Ticket status at the time the sample frame was drawn.	Used to determine pathing at item E24. Only respondents identified by SSA as being Ticket participants, and who indicated that they had never heard of the TTW program, were asked this item.
VRname	State name for Vocational Rehabilitation Agency. Based on state of residence reported at time of survey.	Used at items B29, E28, E30, E32, F2, F6, F8, F10, F20, F29, H7, H12, H16, H18, H21, and H52 to identify, by name, the Vocational Rehabilitation Agency in the respondent's state.

Source: NBS, round 3.

TABLE III.3

ITEMS SKIPPED FOR PROXY RESPONDENTS

Survey Item	Question Text
C18	Taking all things into account, how satisfied are you with your {main/current} job? Would you say very satisfied, somewhat satisfied, not very satisfied, or not at all satisfied?
C39a-C39l	Thinking about your {main/current} job, how much do you agree with each of the following statements? Would you say you strongly agree, agree, disagree, or strongly disagree?
C39a	The pay is good
C39b	The benefits are good
C39c	The {job security is good/work is steady}
C39d	You have a chance for promotion
C39e	You have a chance to develop abilities
C39f	You have recognition or respect from others
C39g	You can work on your own in your job if you want to
C39h	You can work with others in a group or team if you want to
C39i	Your work is interesting or enjoyable
C39j	Your work gives you a feeling of accomplishment or contribution
C39k	Your supervisor is supportive
C39l	Your co-workers are friendly and supportive
H10a-H10d	Now I'm going to read you some statements about the Ticket to Work Program. For each statement, please tell me if it is something you knew before today or not. Is this something you knew before today or not;
H10b	You can, during any month, take back your Ticket and give it to another Employment Network or participating provider.
H10c	To remain in the program, you must participate in the activities described in your individual work plan during the first few years, and work for 3 to 6 months each year during the later years of your participation.
H10d	While you are working, you can keep your Medicare and /or Medicaid benefits.
H11	Before you started participating, how much would you say you knew about the Ticket to Work Program? Would you say a lot, some, a little, or nothing?
H45	Overall, how satisfied are you with the Ticket to Work program? Would you say very satisfied, somewhat satisfied, not very satisfied, or not at all satisfied?
H58	How satisfied are you with how the problem (with the SVR/EN) was solved? Would you say very satisfied, somewhat satisfied, not very satisfied, or not at all satisfied?
H59	Overall, how satisfied are you with the helpfulness of the {State VR/EN} in trying to solve this problem? Would you say very satisfied, somewhat satisfied, not very satisfied, or not at all satisfied?

Source: NBS, round 3.

C. COMPARISONS WITH OTHER QUESTIONNAIRES AND SURVEYS

The NBS contains a number of questions that are found on other survey instruments. Table III.4 provides the names of the studies from which NBS questions have been drawn, their sponsors (where relevant), and the NBS question number. In some instances, the question was asked on multiple studies, in which case all studies are listed.

TABLE III.4

NATIONAL BENEFICIARY QUESTION SOURCES

Study/Source	Sponsor	Question Numbers
A National Study of Health and Activity (NSHA)	Social Security Administration (SSA)	B18, B19, B25a-k, B47a-d, C6, C8, C9, C11, C20a-i, C33a-f, D14, D16-D19, I19, I20, I23, I24, I31, I32, J1, J2, J4-J6, K7, K8a-h
Employment Intervention Demonstration Program (EIDP)	Center for Mental Health Services, Substance Abuse and Mental Health Services Administration (SAMHSA)	B47a-k
State Partnership Initiative Participant Employment Data Form	SSA	C20a-i
Project Network Baseline Survey	SSA	K7, K8a-h
Evaluation of the Effects of the 1996 Welfare Reform Legislation on Children with Disabilities	SSA	E3-10, E12, E13, E15-E19, E20a-d
1996 Survey of Income and Program Participation (SIPP) Wave 5 Functional Limitations and Disability Adult Topical Module	Demographic Survey Division, US Census Bureau	I17, I18, I21, I22, I25, I26, I29, I30, I33-I39, I41, I43, I45-I52, I55-I61
Office of Management and Budget (OMB) Standards for Maintaining, Collecting and Presenting Federal Data on Race and Ethnicity		L1-L2

Source: NBS, round 3.

D. SPECIAL DESIGN CONSIDERATIONS

The NBS survey population represented a wide range of disabilities with varying degrees of severity; in addition, some sample members had several disabling conditions. While the survey could not be designed to overcome all possible challenges, the instrumentation procedures attempted to address three broad categories of common challenges: (1) communication, (2) stamina, and (3) cognitive barriers. Communication challenges include both hearing and speech impairments. "Stamina challenges" include physical and mental fatigue. Cognitive challenges include, but are not limited to, emotional disturbance, difficulty processing questions and responses, lack of complete or specific knowledge, and confusion about the purpose of the interview (Mitchell et al. 2004).

The NBS survey featured several design techniques designed to overcome these challenges. The interviews could be conducted via text typewriter (TTY), Telecommunications Relay Service (TRS), or instant messaging so that persons with severe hearing or speech impairments could be interviewed by telephone. In addition, to maximize survey participation, in-person interviewers obtained the services of sign language translators and made a range of other accommodations when interviewing persons with hearing impairments in their home.

Structured probes were included in the survey instrument, which allowed questions to be rephrased and concepts defined in a standard manner in the event that respondents required clarification or additional information. Additionally, to minimize item nonresponse, the survey instrument included follow-up questions for continuous variables. For example, if a respondent could not provide an exact amount, the "Don't know" response was followed with a modified version of the question that offered response categories. The upper and lower bounds of each category were based on ranges specified by analysts.

All respondents were notified in the introduction to the study that if they began to feel tired the interviewer could stop and the interview could be completed at a later time. Interviewers were also trained to check with respondents about their level of fatigue during the interview. If they sensed that a respondent was tiring, they repeated this and asked the respondent if he or she was OK to continue. The instrument was set up so that the interview could be broken off at any time and a call-back time scheduled. In round 3, 996 cases (about 11 percent of the total sample) were broken off after the interview began (that is, after the screener and cognitive items had been administered and the respondent was in the body of the questionnaire). Of these, 848 cases were later completed (85 percent); 148 were not completed (15 percent).

In general, we attempted to word survey questions simply, clearly, and briefly as well as in an unbiased manner so that respondents could readily understand key terms and concepts. Given the intent of the questions, response categories were appropriate, mutually exclusive, and reasonably exhaustive.

E. CHANGES MADE TO SURVEY INSTRUMENT AT ROUND 3

Some changes were made to the survey instrument at round 3 to update it for administration in 2005, including: (1) changing reference periods from 2004 to 2005, (2) making revisions to accommodate longitudinal respondents who had completed a round 1 interview but did not complete a round 2 interview, (3) adding new items designed to gather more in-depth data from respondents who were not working at the time of the interview, and (4) adding additional consistency checks. These changes are briefly described below.

1. Changes to Reference Periods

Questions that referenced calendar year 2004 during the round 2 survey were changed to 2005. Items affected included those asking about jobs held in 2004, Ticket use in 2004, services

received in 2004, TTW participation in 2004, and insurance coverage in 2004. In a few cases, response categories were also edited to reflect the new data collection period.

2. Adding New Items

At SSA's request, questions were added to section B to gather in-depth information about respondents who were not currently working. For those looking for work, questions about desired work hours, whether a job had been offered in the past four weeks, and why the job was turned down or why the respondent had been unable to find a job (if applicable) were added. Questions were also added asking beneficiaries if they were reluctant to work because their earnings would cause them to lose needed benefits and how much money they would need to make to take a job if they were not currently working. Other items in section B were moved from their round 2 location to accommodate these changes.

In a section C, questions were added to help determine if beneficiaries work fewer hours than they could have because they fear losing benefits. Questions were added to section D to assess the impact of loss or reduction of benefits as a result of work in the prior year. These questions were similar to questions included at prior rounds in section C.

Two questions were added at round 3 to assess the impact of hurricanes Katrina and Rita on employment and housing during the data collection period. Finally, an item was added to assess respondent satisfaction with the survey process. A list of specific items added at round 3 is included in Appendix B.

3. Changes Made to Accommodate the Longitudinal Participant Sample

Revisions to questionnaire pathing and question wording were made to accommodate longitudinal cases that responded to the survey in round 1 but did not complete round 2. At round 3, some questions were skipped for longitudinal respondents who completed one or more prior

rounds, but in other cases, particularly in section E, longitudinal respondents skipped some questions only if they completed round 2. These revisions are summarized in Appendix C.

4. Consistency Checks

Additional consistency checks were built into the round 3 questionnaire to improve data quality. A check was added to calculate and compare monthly income and pre-tax and take home pay as reported in section C (current employment) to earnings and monthly pay reported in section K (income) for all jobs. Consistency checks were also added to sections C, D, and K to flag cases for which reported take-home pay was greater than pre-tax pay or for which the difference between pre-tax and take home pay was greater than 30 percent.

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

IV. DATA COLLECTION

The NBS was executed as a dual-mode survey. Initial interview attempts were made using computer-assisted telephone interviewing (CATI) followed by computer-assisted personal interviewing (CAPI) of nonrespondents. CAPI interviews were attempted with respondents who requested an in person interview, those who needed an in-person interview to accommodate a disability, and those did not have telephones or whose telephone number could not be located. If a sample person was not able to participate in the survey due to his or her disability, a proxy respondent was sought. If no proxy was available and an in-person interview was not possible, the final status of the case was classified as a nonresponse. Sample persons or proxies who requested an in-person interview and who were eligible for field follow-up were held for the start of CAPI data collection.

CATI data collection began in February 2006[15]. In person locating and interviewing of telephone nonrespondents and beneficiaries who requested an in-person interview began in May 2006 and continued, concurrent with CATI interviewing, through September 2006. In total 6,605 cases were completed (including 16 partially completed interviews)—2,508 from the Representative Beneficiary Sample and 3,115 from the Phase 2 and Phase 3 Cross-Sectional Samples. An additional 982 cases were not eligible for the Phase 1 Cross-Sectional Sample, but are included on the data file as completed interviews from the Longitudinal Ticket Participant Sample (for a total of 4,097 Ticket Participant complete interviews)[16].

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

[16] 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 of the interview).

MPR conducted a CATI pretest in December 2003 to test the programmed instrument prior to fielding. Overall, 74 pretest interviews were completed--thirty-two with participants and 42 with nonparticipants. As a result of the pretest, minor instrument changes were identified and programming problems corrected prior to full-scale CATI interviewing. More details of the pretest can be found in the NBS round 1 User's Guide (Wright et al 2009).

A. DATA COLLECTION PROCEDURES

1. Advance Contacts

In an effort to increase respondent trust and rapport, all sample members for whom MPR had a valid address were sent an advance letter and a list of frequently asked questions and answers before the start of data collection. The advance letter, printed on SSA letterhead and signed by an SSA official, identified SSA as the sponsor of the survey and MPR as the survey contractor, explained the purpose of the survey, offered assurances of confidentiality, described the voluntary nature of participation, and included a toll-free number, a TTY number, and an e-mail address for respondents to use to contact MPR with questions or to complete the interview at their convenience. To encourage participation and show appreciation for response, a post-paid incentive payment of $10 was offered to respondents who completed the survey. The advance letters indicated that the interview could be conducted in the sample person's home if he or she was unable to respond by telephone because of a disability. Longitudinal sample members were sent a version of the letter reminding them that they had been contacted the previous year regarding the study and letting them know that we would like to talk to them again in 2006.

In an additional effort to help establish legitimacy, SSA posted information about the survey on the agency Web site and circulated information describing the survey to SSA field offices and the SSA teleservice (800) center. Field offices and the SSA teleservice (800) center were also sent the names of telephone and in-person interviewers involved in the NBS so that these

individuals could be identified as legitimate contacts. If upon receipt of the advance letter, disability beneficiaries contacted their local field office or the SSA 800 number with questions about the survey or its legitimacy, SSA staff could then assure beneficiaries of the study's legitimacy and encourage them to participate.

2. Interviewer Training

CATI interviewers received 14 hours of training over four sessions in February 2006. The CAPI interviewers were trained in three separate 24-hour trainings with each training split across three days. The NBS training included providing interviewers with the background and purpose of the study, a question-by-question review of the instrument, contact protocols, refusal avoidance strategies, and a series of practice interviews. In addition, sensitivity training was included, emphasizing the importance of patience, professionalism, and showing unconditional positive regard for respondents regardless of their impairments. Trainers stressed that the greatest barriers faced by people with disabilities are often others' prejudgments and erroneous images of them. Interviewers were trained to use positive rather than patronizing language and were encouraged to focus on the individual first and the disability last.

To overcome stamina challenges, interviewers were trained to be aware of behaviors that might indicate that a respondent was too fatigued to continue. If a respondent seemed tired, agitated, or distracted, for example, interviewers were encouraged to ask whether the respondent needed to take a break and schedule another time to continue and to set appointments for times when the respondent was most alert. To overcome cognitive challenges, the training focused on neutral, nondirected probing methods (repeating the question, repeating the response categories, asking for more information, stressing generality, stressing subjectivity, and zeroing in) and using active listening skills and patience. Interviewers were instructed to provide neutral feedback and encouragement throughout the survey. They were trained to help keep the

respondent free of distractions, to say the respondent‘s name often, and to avoid using an exaggerated inflection or tone of voice.

3. Locating

Sample member contact information was provided by SSA from administrative records. Prior to the mailing of the advance materials, all addresses were verified or updated using a commercially available database. Over the course of the round 3 data collection, 40 percent of telephone numbers initially provided were identified as invalid and were sent to central office locating. MPR used a variety of techniques for locating updated information, including database searches, calling relatives and friends, receiving updated contact information from SSA, and making in-person visits for field locating. Due to these efforts, approximately 93 percent of the sample was eventually located for interviewing or determined to be ineligible. Of the located sample, 70 percent completed the interview.

4. CATI Data Collection

CATI data collection began in February 2006. In total, 5,104 cases were completed by telephone (77 percent of completes). Sixty-nine percent of the Representative Beneficiary Sample completes (n=1,737) and 82 percent of the Ticket Participant Sample completes (n=3,367) were completed via CATI. Approximately 57 percent of the total completes were obtained before the start of CAPI data collection (May 2006). The telephone survey took, on average, 50 minutes to administer. The interview length ranged from 16 to 180 minutes excluding TTY, TRS, and instant messaging interviews.

a. Assistive Technologies

Several technologies were available to assist with interviewing deaf and hard-of-hearing sample persons by telephone including phone amplifying volume controls, an in-house TTY

machine, TRS, and instant messaging. To minimize respondent burden when using TTY, TRS, and instant messaging, an electronic version of the instrument was maintained that included standard TTY abbreviations and punctuation [such as ~~–g~~a" (go ahead), ~~–n~~u" (number), ~~–o~~ic" (oh, I see)], which interviewers could use to ~~–c~~ut" the question text from the electronic file and ~~–p~~aste" into the TTY text box or instant messaging screen to ask a question. Respondents' answers were then entered into the computerized survey instrument on a second PC. Despite these efforts, the average length of a TTY interview was considerably longer than that of a non-TTY interview. For round 3 of the NBS, the average time to complete a TTY interview was approximately 3 hours and often required several sessions. The shortest TTY interview was about one and a half hours and the longest was 6 hours over several sessions. Interviews conducted by instant messaging were generally about one-third shorter than those using TTY.

For round 3, we identified 94 respondents who were hearing impaired and could potentially be interviewed using TTY, TRS, or instant messaging. In 32 cases, the sample member completed the interview: 6 by TTY, 3 by TRS, 7 by instant messaging, 3 by video relay, 2 using a sign language interpreter, and 11 with an in-person interviewer. An additional 13 cases were completed by proxy. The remainder either did not answer the TTY call or refused, usually citing length of the interview.

5. CAPI Data Collection

In-person interviewing was employed to maximize access to the survey among persons with disabilities. In-person administration can facilitate interviewing of persons with hearing and speech limitations who are unable to participate by telephone, can permit in-person assistance to persons with cognitive challenges, and can improve the locating rate through in-field searching (Mitchell et al. 2004). To control costs, attempts were first made to contact and interview sample

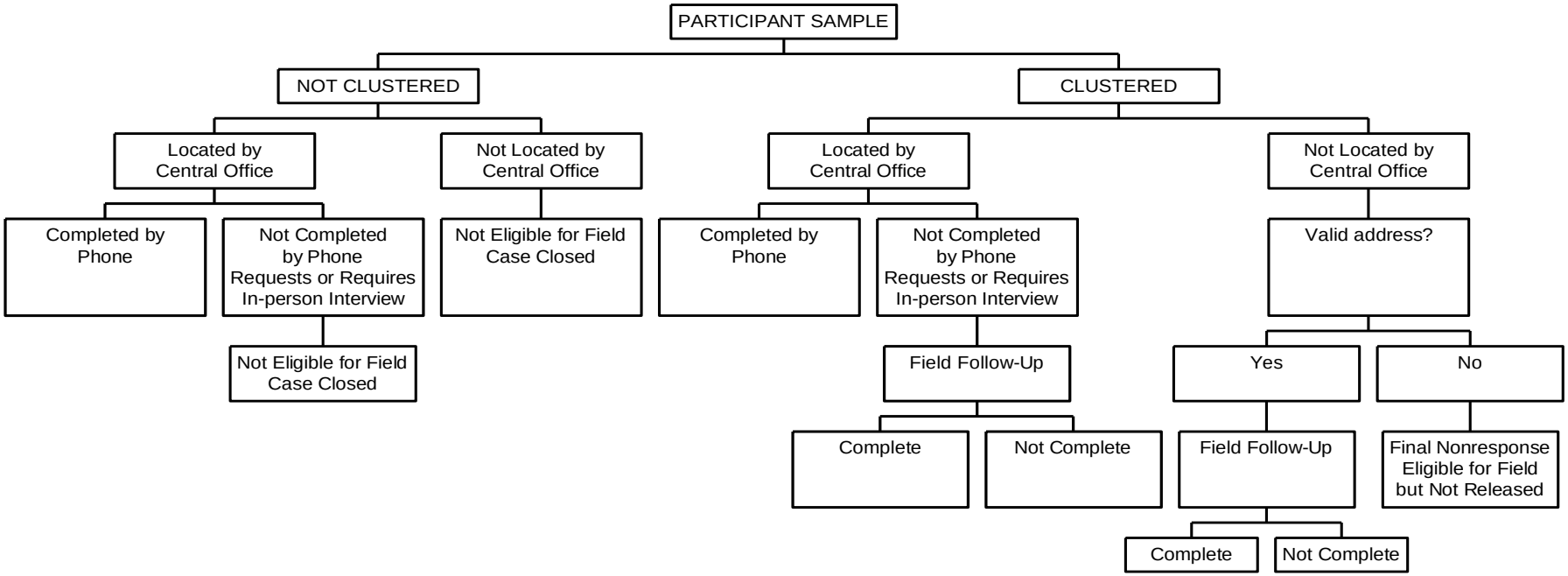
persons via telephone. CAPI interviews were then offered to anyone who requested an in- person interview or who needed an in-person interview to accommodate a disability.

All cases referred for in-person interviewing (refusals, those who were evasive to telephone attempts, and those who requested an in-person interview) were first sent to central office locating. Locating verified or updated the telephone and address if possible before the case was assigned to a field interviewer and provided the field interviewer with a listing of previous addresses. Additionally, cases for which a telephone number could not be located were flagged for CAPI follow-up. As discussed in Chapter II, the unlocated, unclustered outcome-only Ticket Participant Sample was not eligible for CAPI field treatment. For the purpose of data collection, clustered and unclustered cases were subjected to identical predetermined central office locating procedures. Once central office locating was exhausted, clustered cases were sent to the field for in-person locating and unclustered cases were put on hold and received no further locating treatment. See Figure IV.1 for a summary of the CAPI Ticket Participant Sample administration. In all, 2,734 cases, or approximately 30 percent of the total sample, were sent to CAPI interviewers to be conducted in-person. Of these, 61 percent were completed; 185 (7 percent) via CATI, and 1,490 (54 percent) by field interviewers. Field interviewers were trained to encourage sample persons to call in and complete the survey by telephone once they were located to save on data collection costs. Thirty-one percent of the Representative Beneficiary Sample completes (n=771 and 18 percent of Ticket Participant completes (n=730) were obtained via CAPI.

Most cases that were sent to the field (57 percent) were sent because they could not be located or did not have a telephone. Another 22 percent were sent to the field because the sample person initially refused a CATI interview. An additional 18 percent were sent to the field because they were difficult to contact via telephone or had evaded contact efforts. The remaining one percent of cases were sent to the field because they requested an in-person interview.

FIGURE IV.1

NATIONAL BENEFICIARY SURVEY - SAMPLE ADMINISTRATION



To ensure that the highest-quality CAPI data were collected, several Quality Assurance (QA) procedures were in place. Early CAPI data were reviewed for the frequency of item nonresponse and other data problems. Using this information, feedback and additional instruction were given to interviewers who needed it. Second, interview length was checked for patterns of especially lengthy or short interviews, since consistently short or long interviews might indicate data forgery or other problems. Finally, 10 percent of each interviewer‘s cases were randomly selected and verified by either telephone or mail. During the verification, respondents were asked how long the interview lasted, whether the interviewer used a laptop, and what types of questions were asked. In addition, some questions were re-asked to ensure that the answers are the same as those recorded during the interview.

6. Assisted Interviews and Proxy Respondents

To increase opportunities for self-response, ―assisted‖ interviews were also permitted. These interviews were different from proxy interviews because beneficiaries answered most questions themselves. The assistant, typically a family member, provided encouragement, interpretation, and verified answers when needed. In the NBS, we allowed assisted interviews in order to minimize item nonresponse, improve the accuracy of responses, and overcome less limiting conditions (such as difficulty hearing) and language barriers. In all, 238 assisted interviews were conducted (approximately 4 percent of all completes) during round 3.

As a last resort, proxy respondents were used to complete the survey on behalf of respondents who could not complete the survey themselves (even with assistance) either by telephone or in-person. This included sample persons with severe communication impairments, those with severe physical disabilities that precluded participation (in any mode), and those with mental impairments that might have compromised data quality. Using the beneficiary instead of a proxy when possible was strongly favored because sample members generally provide more

complete and more accurate information than proxy respondents. However, allowing the use of proxies when necessary minimized the risk of nonresponse bias that would have resulted from the exclusion of individuals with severe physical or cognitive impairments.

In the NBS, we used an innovative ―mini-cognitive test" designed expressly for the survey to identify when proxy respondents were needed.[17] The screener provided interviewers with a tool for evaluating when to seek a proxy rather than leave the decision to their discretion or to gatekeeper advice. The test combined the ability to understand the survey topics with elements of informed consent.

Specifically, we asked three questions at the start of the interview. First, we gave a general description of the survey topics to be covered (your health, daily activities, and any jobs you might have) and asked the respondent to state the topics in his or her own words. Second, we described the voluntary nature of the survey and asked respondents to state, in their own words, what that description meant to them. Third, we described the confidential nature of the respondents' answers and asked them to state what that description meant. If respondents were unable to restate accurately any description after two attempts, we asked if someone else could answer questions on their behalf.

For cases in which a sample person or knowledgeable informant expressed that a proxy would be necessary, several guidelines were used to determine whether a proxy would be appropriate. These guidelines included using proxies only when the sample member's physical or mental condition precluded self-response, selecting the most knowledgeable proxy, and ensuring that the proxy answered on behalf of the sampled respondent rather than offering his or

[17] Westat designed the test as part of the design of the Ticket to Work evaluation; MPR modified it after pretesting.

her own opinions. Interviewers were trained to overcome gatekeepers‘ objections, and to give sample members the opportunity to speak for themselves whenever possible.

At round 3, proxy interviews were completed for 1,286 sample persons (19 percent of all completes). In most cases (approximately 76 percent), a proxy was necessary because the sample person failed the cognitive assessment or was otherwise determined to be unable to respond due to a cognitive or mental impairment. Interviews were completed by proxy for 637 sample persons in the Representative Beneficiary Sample (25 percent of completes) and 503 sample persons in the Ticket Participant Sample (12 percent of completes).

B. CASE DISPOSITION SUMMARIES

A total of 2,508 cases from the Representative Beneficiary Sample and 4,097 cases from the Ticket Participant Sample Cross-Sectional Sample were completed; 215 beneficiaries and 46 TTW participants were determined to be ineligible for the survey. Ineligible cases included sample persons who were deceased, no longer living in the continental United States, who were incarcerated or institutionalized, or who were denied benefits since the time of sample selection or who had never received SSA benefits. An additional 982 Ticket Participant Phase 2 Longitudinal cases were completed that were not eligible for the round 3 cross-sectional sample (for a total of 4,097 Ticket Participant completes in all). Table IV.1 provides a summary of final case disposition for all released cases in the cross-sectional sample by sampling strata.

TABLE IV.1

SUMMARY CASE DISPOSITION BY SAMPLE TYPE AND SAMPLING STRATA FOR CROSS-SECTIONAL CASES

		Complete			Ineligible			Refused			Unlocated			Non-Respondents		
	Total Sample	Count	Un-weighted Percent	Weighted Percent	Count	Un-weighted Percent	Weighted Percent	Count	Un-weighted Percent	Weighted Percent	Count	Un-weighted Percent	Weighted Percent	Count	Un-weighted Percent	Weighted Percent
National Representative Beneficiary Sample																
Age 18-29	943	698	74.0	74.0	77	8.2	8.2	62	6.6	6.6	69	7.3	7.3	37	3.9	3.9
Age 30-39	941	672	71.4	71.4	56	6.0	6.0	101	10.7	10.7	63	6.7	6.7	49	5.2	5.2
Age 40-49	935	711	76.0	76.0	49	5.2	5.2	89	9.5	9.5	41	4.4	4.4	45	4.8	4.8
Age 50-64	563	427	75.8	75.8	33	5.9	5.9	59	10.5	10.5	23	4.1	4.1	21	3.7	3.7
Total Beneficiary	3,382	2,508	74.2	75.2	215	6.4	6.0	311	9.2	9.9	196	5.8	4.8	152	4.5	4.2
Cross-Sectional Ticket Participant Sample																
Phase 2																
Traditional	867	733	84.5	84.6	9	1.1	1.1	69	8.0	7.8	27	3.1	3.0	29	3.3	3.5
Milestone-Outcome	937	663	70.8	78.8	11	1.2	1.0	85	9.1	10.9	17	1.8	1.8	65	6.9	7.5
Outcome-Only	993	666	67.1	70.9	9	.9	3.2	107	10.8	13.6	13	1.3	4.2	75	7.6	8.1
Total Phase 2	2,797	2,062	73.7	83.4	29	1.0	1.1	261	9.3	8.4	57	2.0	2.9	169	6.0	4.2
Phase 3																
Traditional	444	369	83.1	83.1	6	1.4	1.3	42	9.5	9.4	13	2.9	2.9	14	3.2	3.2
Milestone-Outcome	444	362	81.5	82.9	7	1.6	1.4	38	8.6	7.8	18	4.1	4.1	19	4.3	3.9
Outcome-Only	485	322	66.4	73.1	4	0.8	0.9	55	11.3	11.9	18	3.7	5.1	28	5.8	8.9
Total Phase 3	1,373	1,053	76.7	82.8	17	1.2	1.3	135	9.8	9.3	49	3.6	3.2	61	4.4	3.4
Total Participant	4,170[a]	3,115	74.7	83.1	46	1.1	1.2	396	9.5	8.8	106	2.5	3.0	230	5.5	3.8
Combined Sample																
Total Sample**	7,552	5,623	74.5	75.2	261	3.5	5.9	707	9.4	9.9	302	4.0	4.8	382	5.1	4.2

Source: NBS, round 3.

Note: The number of completed cases includes 16 partially completed interviews: 10 in the Ticket Participant Sample and 6 in the Representative Beneficiary Sample.

[a]The total number of Ticket Participant cases in the Cross-Sectional Sample does not include 982 cases from the longitudinal sample that were not eligible for the Cross Section (were not TTW Participants at Round 3).

** The weighted percentages can be calculated as a weighted average of the Representative Beneficiary and Ticket Participant Samples. This average is dominated by the percentages from the Representative Beneficiary Sample.

C. LONGITUDINAL RESPONSE

At round 3, interviews were attempted with all sample members selected into the Phase 1 Ticket Participant Longitudinal Sample, whether or not they had completed an interview at round 1, and all sample members selected into the Phase 2 Ticket Participant Longitudinal Sample, whether or not they had completed an interview at round 2. Interviews were completed with 953 Phase 1 longitudinal cases (65 percent of the total sample) for a weighted response rate of 73.5 percent, and 969 Phase 2 longitudinal cases (71 percent of the total sample), for a weighted response rate of 83.2.

Of the Phase 1 longitudinal cases, 767 (52 percent of the total sample) completed all three survey rounds; 116 completed round 3 but did not complete round 1, and 108 completed round 3 but did not complete round 2 (see Table IV.2). Of the Phase 2 longitudinal cases, 831 completed both a round 2 and round 3 interview (62 percent of the total sample; see Table IV.3).

TABLE IV.2

HISTORY OF CASE DISPOSITION FOR PHASE 1 LONGITUDINAL ROUND 3 COMPLETES

	Round 2 Complete		Round 2 Refusal		Round 2 Ineligible		Round 2 Unlocated		Round 2 Nonresponse		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Round 1 Complete	767	91.6	15	1.8	0	0.0	22	2.6	33	3.9	837	87.8
Round 1 Refusal	19	73.1	3	11.5	0	0.0	3	11.5	1	3.9	26	2.7
Round 1 Ineligible	4	50.0	0	0.0	4	50.0	0	0.0	0	0.0	8	0.8
Round 1 Unlocated	23	67.7	0	0.0	0	0.0	9	26.5	2	5.9	48	5.0
Round 1 Nonresponse	32	66.7	1	2.1	0	0.0	2	4.2	13	27.1	34	3.6
Total	**845**	**88.7**	**19**	**2.0**	**4**	**0.4**	**36**	**3.8**	**49**	**5.1**	**953**	**100.0**

Source: NBS round 3.

TABLE IV.3

ROUND 2 CASE DISPOSITION FOR PHASE 2 LONGITUDINAL ROUND 3 COMPLETES

Round 2 Complete		Round 2 Refusal		Round 2 Ineligible		Round 2 Unlocated		Round 2 Nonresponse		Total	
Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
831	85.8	33	3.41	1	0.1	38	3.9	66	6.8	969	100

Source: NBS round 3.

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

V. VARIABLE CONSTRUCTION AND EDITING

The NBS data files contain several types of variables: unedited and edited questionnaire variables, imputed variables and imputation flags, coded verbatim responses, variables masked for the Public Use File, constructed variables derived from questionnaire variables, weights, survey administration variables, and SSA administrative data.[18] This chapter provides an overview of the types of variables on the file and variable naming conventions as well as additional details on coded items and select constructed variables.

A. EDITING OF QUESTIONNAIRE VARIABLES

Questionnaire variables are survey items collected directly from the respondent. On the NBS data files, these variables are distinguished by a two-part name with the first part of the variable name representing the section of the questionnaire where the question originates and the second part of the variable name representing the numerical question from the questionnaire (for example, question F11 comes from section F of the questionnaire and is question 11). Variables on the file are also preceded by an R3_ to identify them as round 3 variables.

The NBS data were thoroughly reviewed for discrepancies that might have resulted from programming or interviewer errors. Editing was performed to resolve any inconsistencies in skip patterns. Editing also included a review and resolution of some outlier values by recoding either to an appropriate valid value or to a value of missing (.D=don‘t know). For key variables, these responses were imputed along with other missing values. In consultation with SSA and research analysts, we took the general approach of editing only those cases where there appeared to be an

[18] In general, unedited variables are those which contain the original response to a single questionnaire item.

obvious data entry or respondent error. As a result, while a substantial amount of time was spent meticulously reviewing individual responses, some suspect values remain on the file. For more information on data problems and the completeness of the survey data set, see the –National Beneficiary Survey: Round 3 Data Cleaning and Identification of Data Problems Report" (Wright and Barrett 2008).

B. IMPUTATION OF MISSING VALUES

A case may be missing data for a particular item because the item was logically skipped (the respondent was not eligible to receive the item), the respondent refused the item or gave a –don't know" response, there was an interviewer or programming error that resulted in a loss of data, or the case was a partial complete and is missing data for some items in the survey. Data for cases completed up through H61 (or G61 if the respondent was not eligible for section H) were included on the file as partial completes. All subsequent items for these cases were coded as .P. Table V.1 provides a summary of missing value codes and their description. For NBS, missing data due to don't know or refused responses and those missing because the case was partially completed (.D, .R, and .P) were imputed for selected variables on the file.

TABLE V.1

MISSING VALUES AND DESCRIPTION

Value	Description
L	Logical skip: respondent not eligible to receive the item
D	Don't know: respondent did not know how to answer the item
R	Refused: respondent refused to provide a response to the item
M	Missing data: data are missing due to interviewer or programming error
P	Partial complete: data are missing due to partial interview

Source: NBS, round 3.

Variables were selected for imputation based on their level of missing data and their analytic importance. Variables imputed included those related to race and ethnicity, disability status, current employment, health, income, and personal and household characteristics. A complete list of variables selected for imputation and the specific imputation procedures employed for each item can be found in Chapter VII. Imputed variables share the same name as the original variable but end in an ―i". The original non-imputed variables are retained on the Restricted Access File. Imputation flags are also included on the Restricted Use File and indicate that a case has been imputed and describes the method of imputation (see Table V.2). At round 3, the flag ―7=Longitudinal Imputation" was added indicating that for some variables, the imputed value from round 1 was used. Imputation flag variables share the same name as the original variable and end in ―iflag" (for example, ―BMI_cat_i" is the imputed version of the constructed variable C_BMI. BMI_cat_iflag indicates which cases were imputed and the method used for that imputation).

TABLE V.2

IMPUTATION FLAG VALUES AND DESCRIPTION

Imputation Flag Value	Description
0	No Change
1	Logical Imputation
2	Administrative Data
3	Hotdeck Imputation
4	Imputed by Distributional Assumptions
5	Imputed by Descriptive Statistic
6	Constructed from Imputed variables
7	Longitudinal Imputation
L	Logical Skip
P	Partial

Source: NBS, round 3.

C. CODING OF VERBATIM RESPONSES

The NBS questionnaire includes a number of questions designed to elicit open-ended responses. To make it easier to use the data connected with these responses in an analysis, we grouped the responses and assigned them numeric codes when possible. The methodology used to code each variable depended upon the content of the variable. Three kinds of questions (described below) on the NBS did not have designated response categories; rather, the response to these questions was recorded verbatim:

1. ***Open-ended questions*** have no response options specified (such as E43—Why are you no longer receiving services from your employment network?). For these items, interviewers recorded the verbatim response. Using common responses, we developed categories and reviewed them with analysts. Coders then attempted to code the verbatim response into an established category. If the response did not fit into one of those categories, it was coded as —other."
2. ***"Other/specify"*** is a response option for questions that have a finite number of possible answers that may not necessarily capture *all* possible responses. A good example is: "Did you do anything else to look for work in the last four weeks that I didn't mention?" For questions of this type, respondents are asked to specify an answer to the question "anything else?" or "anyone else?"
3. ***Field-coded responses*** are answers coded by interviewers into a predefined response category without reading the categories aloud to the respondent. If none of the response options seem to apply, interviewers select an "other specify" category and type in the response.

As part of data processing at round 1, we examined a portion of all verbatim responses in an attempt to uncover dominant themes for each question. Based on this initial review, we developed a list of categories and decision rules for coding verbatim responses to open-ended items. In addition, supplemental response categories were added to some field-coded or other/specify items to facilitate coding if there were enough such responses and they could not be back-coded into pre-existing categories. (A list of all open-ended items assigned additional categories during the coding process appears in Appendix D.) 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 and round 2 coding were applied at round 3. Additional categories were added at round 3 for a small number of items if there were a significant number of common responses that did not fit into previously developed categories. If during coding, it became apparent that changes to the coding scheme were necessary (for example adding additional categories or clarifying coding decisions), new decision rules were discussed and documented. Verbatim responses were sorted alphabetically by item for coders and could be filtered by coding status so that new decision rules could be easily applied to cases that had been previously coded. When it was impossible to code a response, when responses were invalid, or when they could not be coded into a given category, we assigned a two-digit supplemental code to the response (see Table V.3). The verbatim responses themselves are excluded from the data files. (See Barrett and Wright (2008) for full details regarding the back-coding procedures.)

TABLE V.3

SUPPLEMENTAL CODES FOR OTHER/SPECIFY CODING

Code	Label	Description
94	Invalid Response	Indicates that the response should be deleted.
95	Refused	The verbatim indicates the respondent refused to answer the question.
96	Duplicate Response	The verbatim response has already been selected in a ‗code all that apply' item.
98	Don't Know	The verbatim indicates the respondent did not know the answer.
99	Not Codeable	Indicates that a code cannot be assigned based on the verbatim response.

Source: NBS, round 3.

Two special cases of verbatim response coding are discussed in more detail below: health condition, and industry and occupation coding.

1. Health Condition Coding

Responses to questions on health conditions required a specific type of open-ended coding. In Section B of the questionnaire, each respondent was asked to cite the main and secondary physical or mental conditions that limit the kind or amount of work or daily activities he or she can do. Main conditions could be reported as one of four items: B2 (main reason limited), B6 (main reason eligible for benefits), B12 (main reason was eligible for benefits if not currently eligible), and B15 (main reason limited when first started getting disability benefits). The main purpose of items B6, B12, and B15 was to collect information on a health condition from people who reported no limiting conditions in B2. For example, if respondents said that they had no limiting conditions, they were asked if they were currently receiving benefits from Social Security. If they answered ―yes," they were asked for the main reason that made them eligible for benefits (B6). If respondents said that they were not currently receiving benefits, they were asked whether they had received disability benefits in the last five years. If they answered ―yes," they were asked for the condition that made them eligible for Social Security benefits (B12), or for the reason that first made them eligible if they no longer had that condition (B15). If respondents said that they had not received disability benefits in the last five years, they were screened out of the survey and coded as ineligible. Each response to B2, B6, B12, and B15 was assigned a value for the three constructs. Although respondents were asked to cite one ―main" condition in B2, B6, B12, or B15, many listed more than one. These additional responses were maintained under the main condition variable and coded in the order in which they were recorded. Longitudinal cases that completed rounds 1 or 2 skipped items B6, B12, and B15 at round 3.

For each item on a main condition, respondents were also asked to list any other, or secondary, conditions. For example, respondents reporting a main condition at B2 were asked at

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

As in rounds 1 and 2, the respondents‘ verbatim responses were coded according to the International Classification of Diseases, 9th revision, Clinical Modification (ICD-9-CM) five-digit coding scheme. The ICD-9 is a classification of morbidity and mortality information that was developed in 1950 to index hospital records by disease for data storage and retrieval. The ICD-9 was available in hard copy for each of the coders. Coders, many of whom had previous medical coding experience, attended an eight-hour training session before coding and were instructed to code to the highest level of specificity possible. Responses that were not specific enough for a five-digit code were coded to four (subcategory) or three digits (category codes). Responses that were not specific enough for even three- or four-digit ICD-9 codes were coded either as a physical problem (not specified) or to broader categories representing disease groups. (See Table V.4 for a list of the broad categorical and supplementary codes.) Although respondents were asked to cite one ―main‖ condition in B2, B6, B12, or B15, many listed more than one. In cases in which multiple, distinct conditions were provided by the respondent, all conditions were coded (for instance, three distinct conditions would be recorded and coded as B2_1, B2_2, and B2_3).

TABLE V.4

BODY SYSTEM DIAGNOSIS GROUPS
(C_MAINCONBODYGROUP_1-_9, C_SECCONBODYGROUP_1-_9, C_REASBECELIGBODYGROUP)

Code	Label	Description of ICD-9 codes	Corresponding ICD-9 codes
00	Other	Other and unspecified infectious and parasitic disease; alcohol dependence syndrome and drug dependence; learning disorders and developmental speech or language disorders; complications of medical care, not elsewhere classified	136.0-136.9, 303.00-304.90, 315.00-315.39, 999.0-999.9
01	Infectious and parasitic diseases	Borne by a bacterium or parasite and viruses that can be passed from one human to another or from an animal/insect to a human including tuberculosis, HIV, other viral diseases, and venereal diseases (excluding other and unspecified infectious and parasitic diseases)	001.0-135, 137.0-139.8
02	Neoplasms	New abnormal growth of tissue, i.e., tumors and cancer, including malignant neoplasms, carcinoma in situ, and neoplasm of uncertain behavior	140.0–239.9
03	Endocrine/nutritional disorders	Thyroid disorders, diabetes, abnormal growth disorders, nutritional disorders, and other metabolic and immunity disorders	240.0–279.9
04	Blood/blood-forming	Diseases of blood cells and spleen	280.0–289.9
05	Mental disorders	Psychoses, neurotic and personality disorders, and other non-psychotic mental disorders including mental retardation (excluding alcohol and drug dependence and learning, developmental, speech, or language disorders)	290.0–302.9, 305.00-314.9, 315.4-319
06	Diseases of nervous system	Disorders of brain, spinal cord, central nervous system, peripheral nervous system, and senses including paralytic syndromes, and disorders of eye and ear	320.0-389.9
07	Diseases of circulatory system	Heart disease, disorders of circulation, and diseases of arteries, veins, and capillaries	390-459.9
08	Diseases of respiratory system	Disorders of the nasal, sinus, upper respiratory tract, and lungs including chronic obstructive pulmonary disease	460-519.9
09	Diseases of digestive system	Diseases of the oral cavity, stomach, esophagus, and duodenum	520.0-579.9
10	Diseases of genitourinary system	Diseases of the kidneys, urinary system, genital organs, and breasts	580.0-629.9
11	Complications of pregnancy, child birth, and the puerperium	Complications related to pregnancy or delivery, and complications of the puerperium	630-677

TABLE V.4 *(continued)*

Code	Label	Description of ICD-9 codes	Corresponding ICD-9 codes
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.	18
95	Refused	Verbatim indicates respondent refused to answer the question.	No ICD-9 codes
96	Duplicate condition reported	The condition has already been coded for the respondent.	No ICD-9 codes
97	No condition reported	The verbatim does not contain or symptom to condition to code.	No ICD-9 codes
98	Don‘t know	The respondent reports that he/she does not know the condition.	No ICD-9 codes
99	Uncodeable	A code cannot be assigned based on the verbatim response.	No ICD-9 codes

Source: NBS, round 3.

We ensured that responses were coded according to the proper protocols in several ways. First, we did an initial quality assurance check, per coder, for the first several cases that were coded. In addition, cases were randomly selected during the coding process for supervisor review. In total, approximately 20 percent of all coded responses were reviewed by a supervisor,

including cases flagged by coders for review that they were unable to code or did not know how to code. Approximately 8 percent of all cases were recoded. In the course of this work, additional decision rules were developed to clarify and document the coding protocol. These decisions were discussed with coders and posted to ensure that responses were coded consistently and accurately throughout the coding process. As for other open-ended items, when new decision rules were added, previously coded responses were reviewed and re-coded if necessary.

After the ICD-9 coding was complete, we processed the health condition variables into a series of constructed variables that grouped health conditions into broad disease groups. In addition to the body system classifications represented in Table V.4 (C_MAINCONBODYGROUP_1-_9, C_SECCONBODYGROUP_1-_9), primary diagnosis groups were formed that provide separate categories for HIV/AIDS, schizophrenia, major affective disorders, mental retardation, visual impairments, hearing impairments, and speech disorders (C_MAINCONDIAGGRP_1-_9, C_SECCONDIAGGRP_1-_9; see Table V.5 for codes). Additional constructs collapse these categories into four broad groups and are provided on the Public Use File (C_MAINCONCOLDIAGGRP_1-_9, C_SECCONCOLDIAGGRP_1-_9; see Table V.6 for codes). A set of separate constructs was also created to summarize responses provided at B6, B12, and B15 (C_REASBECELIGICD9, C_REASBECELIGDIAGGRP, C_REASBECELIGCOLDIAGGRP, and C_REASBECELIGBODYGROUP) These constructs clarify the eligibility of sample members who indicate at B1 and B2 that they do not have a disabling condition.

TABLE V.5

PRIMARY DIAGNOSIS GROUPS
(C_MAINCONDIAGGRP_1-_9, C_SECCONDIAGGRP_1-_9, C_REASBECELIGDIAGGRP)

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 pregnancy, childbirth and the puerperium; conditions in the perinatal period; symptoms, signs and ill-defined conditions; complications of medical care, not elsewhere classified; physical problems not elsewhere classified.	136.0-136.9, 303.00-304.93, 315.00-315.39, 630-677, 760.0–779.9, 780.01-784.2, 784.60-799.99, 999.0-999.9, 11,15, 16, 18
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, other viral diseases, and venereal diseases (excluding HIV and other and unspecified infectious and parasitic diseases)	001.0-041.9, 045.00-135, 137.0-139.8, 01
02	HIV/AIDS	HIV infection	042
03	Neoplasms	New abnormal growth of tissue, i.e., tumors and cancer, including malignant neoplasms, carcinoma in situ, and neoplasm of uncertain behavior	140.0–239.9, 02
04	Endocrine/nutritional Disorders	Thyroid disorders, diabetes, abnormal growth disorders, nutritional disorders, and other metabolic and immunity disorders	240.0–279.9, 03
05	Blood/ blood-forming Diseases	Diseases of blood cells and spleen	280.0–289.9, 04
06	Schizophrenia/psychoses	Schizophrenic disorders	295.00-295.95
07	Major affective disorders	Affective psychoses including major depression and bipolar disorder	296.00-296.99
08	Other mental disorders	Organic psychotic conditions, paranoid states, neurotic disorders, personality disorders, and other non-psychotic mental disorders (excluding alcohol and drug dependence and learning /developmental speech or language disorders, schizophrenia, and major affective disorders)	290.0–294.9, 297.0-302.9, 305.00-314.9, 315.4-316, 05
09	Mental retardation	Mild mental retardation and other specified and unspecified mental retardation	317-319
10	Visual impairment	Disorders of the eye and adnexa	360.00-379.99
11	Hearing impairment	Disorders of the ear and mastoid process	380.00-389.9
12	Speech impairment	Asphasia, voice disturbance, other speech disturbance	784.3-784.5
13	Other diseases of nervous system	Disorders of brain, spinal cord, central nervous system, peripheral nervous system, and senses, including paralytic syndromes, excluding disorders of eye and disorders of ear	320.0-359.9, 06

TABLE V.5 *(continued)*

Code	Label	Description of ICD-9 codes	Corresponding ICD-9 codes
14	Diseases of circulatory system	Heart disease, disorders of circulation, and diseases of arteries, veins, and capillaries	390-459.9, 07
15	Diseases of respiratory system	Disorders of the nasal, sinus, upper respiratory tract, and lungs including chronic obstructive pulmonary disease	460-519.9, 08
16	Diseases of digestive system	Diseases of the oral cavity, stomach, esophagus, and duodenum	520.0-579.9, 09
17	Diseases of genitourinary system	Diseases of the kidneys, urinary system, genital organs, and breasts	580.0-629.9, 10
18	Diseases of skin/ subcutaneous tissue	Infections of the skin, inflammatory conditions, and other skin diseases	680.0-709.9, 12
19	Diseases of musculoskeletal system	Muscle, bone, and joint problems including arthropathies, dorsopathies, rheumatism, osteopathies, and acquired musculoskeletal deformities	710.0-739.9, 13
20	Congenital anomalies	Problems arising from abnormal fetal development, including birth defects and genetic abnormalities	740.0-759.9, 14
21	Injury and poisoning	Problems that result from accidents and injuries including fractures, brain injury, and burns (excluding complications of medical care not elsewhere classified)	800.00–998.9, 17
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 symptom or condition to code.	No ICD-9 codes
98	Don‘t know	The respondent reports that he/she does not know the condition.	No ICD-9 codes
99	Uncodeable	A code cannot be assigned based on the verbatim response.	No ICD-9 codes

Source: NBS, round 3.

TABLE V.6

PRIMARY DIAGNOSIS CODES COLLAPSED
(C_MAINCONCOLDIAGGRP_1-_9, C_SECCONCOLDIAGGRP_1-_9, C_REASBECELIGDIAGGRP)

Code	Label	Description of ICD-9 codes	ICD-9 and Two digit codes
00	Other	Infectious and parasitic diseases; neoplasms; endocrine/nutritional disorders; blood/blood-forming diseases; alcohol dependence syndrome and drug dependence; learning disorders and developmental speech or language disorders; disorders of nervous system; disorders of circulatory system; diseases of respiratory system; diseases of digestive system; diseases of genitourinary system; complications of pregnancy, childbirth and the puerperium; diseases of skin/subcutaneous tissue; conditions in the perinatal period; congenital anomalies; symptoms, signs and ill-defined conditions; injury and poisoning; physical problems not elsewhere classified	001.0-139.8, 01, 140.0–239.9, 02, 240.0–279.9, 03, 280.0–289.9, 04 ,303.00-304.93, 315.00-315.39, 320.0-359.9, 06, 390-459.9, 07 460-519.9, 08, 520.0-579.9, 09, 580.0-629.9, 10, 630-677, 11, 680.0-709.9, 12, 740.0-759.9, 14, 760.0–779.9, 15 780.01-784.2, 784.6-799.99, 16, 800.00–999.9, 17, 18
01	Mental Illness	Organic psychotic conditions, paranoid states, other non-organic psychoses, psychoses with origin specific to childhood, neurotic disorders, personality disorders, and other non-psychotic mental disorders (excluding alcohol dependence syndrome and drug dependence; learning disorders and developmental speech or language disorders; and mental retardation)	290.0-316, 05
02	Mental Retardation	Mild mental retardation and other specified and unspecified mental retardation	317-319
03	Muscular/Skeletal	Muscle, bone, and joint problems including arthropathies, dorsopathies, rheumatism, osteopathies, and acquired musculoskeletal deformities	710.0-739.9, 13
04	Sensory Disorders	Visual, hearing, and speech disorders	360.00-389.9, 784.3-784.5
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 symptom or condition to code.	No ICD-9 codes
98	Don‘t know	The respondent reports that he/she does not know the condition.	No ICD-9 codes
99	Uncodeable	A code cannot be assigned based on the verbatim response.	No ICD-9 codes

Source: NBS, round 3.

2. Industry and Occupation

Information about the sample member's current employment and employment in 2003 was collected in section C and section D of the questionnaire. For each job, respondents were asked to list their occupation (C2 and D4) and the type of business or industry (C3 and D5) where they were employed. Verbatim responses to the occupation items were coded using the Bureau of Labor Statistics' 2000 Standard Occupational Classification (SOC).[19] 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 based upon 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 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). See Appendix E for a list of the three-digit minor groups classified by major groups.

Verbatim responses to the industry items were coded using the 2002 North American Industry Classification System (NAICS).[20] The NAICS is an industry classification system that groups establishments into industrial categories based on the activities in which those establishments are primarily engaged. The NAICS uses a hierarchical coding system to classify all economic activity into 20 industry sectors. For the NBS, we coded NAICS industries to three digits: the first two numbers specify industry sector and the third number specifies the sub-

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

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

sector. See Appendix F for a list of the industries and codes. Both the SOC and NAICS coding schemes are used in most federal surveys and provide uniformity and comparability across data sources.

MPR developed supplemental codes for responses to questions about occupation and industry that were not codeable to a three-digit SOC or NAICS code. Table V.7 lists the occupation and industry supplemental codes.

TABLE V.7

SUPPLEMENTAL CODES FOR OCCUPATION AND INDUSTRY CODING

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

Source: NBS, round 3.

In total, approximately 15 percent of all coded responses were reviewed by a supervisor, including cases flagged by coders for review that they were unable to code or did not know how to code. Approximately 4 percent of all cases were recoded.

The verbatim responses provided at C2 and C3 are not included on the data file. The coded responses to C2 for each job listed are found in the constructed variables C_MainCurJobSOC, C_CurJob2SOC- C_CurJob4SOC. The coded responses to C4 are found in C_MainCurJobNAICS, C_CurJob2NAICS- C_CurJob4NAICS.

D. CONSTRUCTED VARIABLES

The NBS data file preparation included creating more than 300 constructed variables in order to simplify the data file and assist the user. Constructed variables are created by combining information from two or more other sources of data to create one variable. The algorithms and specifications used to create the constructed variables are included in the data file codebooks.

Constructed variables are positioned to appear at the end of the section of variables from which they were created. All constructed variables begin with "C_" succeeded by a brief description of what the variable measures (for example, "C_TotCurWkHours" measures the total weekly hours the respondent worked at all of the jobs he/she listed).

For the NBS, the constructed variables fall into several categories, which are briefly described below. A list of constructed variable names and their description can be found in Appendix G.

1. Survey Administration

The first type of constructed variable includes survey administration and respondent descriptor variables. Included in this set of constructed variables are C_Rtype (indicating whether the interview was completed by the sample member or a proxy respondent), C_IntMode (CAPI or CATI interview), C_Resptype (indicating whether the interview was completed by the sample member only, the sample member with help, or a proxy only), and C_Intage (age at interview). In some cases, constructs were based on sampling variables, for example, C_PaymentType (EN payment type), and C_Cohort (sampling cohort). Other variables on the Restricted Access File identify longitudinal respondents: R3_R1long (Phase 1 Longitudinal Sample Member), R3_R2long (Phase 2 Longitudinal Sample Member), and their response status at rounds 1, 2, and 3 (R3_STATUSR1R2R3). These constructs are positioned at the beginning of the file, prior to the questionnaire sections.

2. Logical Zero

To reduce the number of legitimate missing responses originating from survey skip patterns, ―logical zero" constructs were created for variables assessing the amount of income the sample member received from a variety of sources in the month prior to interview (based on K3, K7a-K7h, K12, and K15). These constructs included the amount earned from jobs last month (C_LstMnthPay), the amount received from private disability insurance (C_AmtPrivDis), worker's compensation (C_AmtWorkComp), veteran's benefits (C_AmtVetBen), public assistance (C_AmtPubAssis), unemployment (C_AmtUnemply), private pension (C_AmtPrivPen), food stamps (C_AmtFoodStamp), other government programs (C_AmtOthGov), other sources on a regular basis (C_AmtOthReg), and from other sources on a nonregular basis (C_AmtOthNonReg). For example, if the respondent reported he or she did not receive private disability insurance last month (question K6a), the follow-up question asking how much private disability insurance was received (question K7a) was skipped. During data processing, such .L (logical skip) responses were recoded to $0. Thus, if the sample member reported not receiving private disability insurance the previous month, then the value of C_AmtPrivDis was ―$0." Logical zero constructed variables are identified in the codebook user notes.

3. Duration and Amount Standardization

Throughout the NBS questionnaire, respondents had the option of reporting contacts with providers, income, and expenditures in the unit of their choosing—for instance, daily, weekly, or monthly. The NBS questionnaire was designed with the expectation that allowing respondents to select the time frame (ideally, the time frame with which they were most comfortable) would improve data quality. In these situations, the amount and the unit reported by the respondent existed as two distinct variables in the survey data. For example, question C12amt asked for the

amount paid on a job and C12hop, how often the amount was paid. To aid the user, constructed variables were created to standardize the time frame, resulting in a single variable (for example, C_MainJobHrPay) in one unit. In section C and D both hourly pay (C_MainCurJobHrPay, C_MainJobHrPay2004) and monthly pay variables were created (C_MainCurJobMnthPay,C_MainCurJobMnthPayTH,C_MainJobMnthPay2004,C_MainJobMnthPayTH2004). Time to report one's current job to SSA was standardized to a week unit (C_MainCurJobRepSSA). Household income, as reported at L23Aamt and L23Ahop was standardized to an annual unit (C_HhInc2004). Variables in section G referencing cost of services (C_ServCost2004, C_TotSerCost2004), costs of equipment and personal assistance services (C_CurMnthEquipExp, C_CurMnthPASExp, and C_TotCurEquipPASExp), duration of visits with provider (C_DurProvVisit), number of contacts with provider (C_NumProvCont), and total money received from ENs (C_TotMoneyENS2004) were also created to standardize reporting units. The NBS codebook provides the specifications used to create these variables in the construct specification notes for each variable.

4. Pathing Combinations

Other constructs were created to combine or summarize survey responses when answers could be provided in multiple places. For example, respondents could report current Medicare coverage in J1 when explicitly probed for this type of insurance and also at J9 ("What kinds of health insurance coverage do you have?") if they reported having no current insurance at J1-J5. In this case, a construct was created that checked both J1 and J9 to determine if the respondent indicated Medicare coverage at either item (C_CurMedicare). This type of construct was created for all health insurance variables in section J. Similar constructs were created to flag awareness of the Ticket to Work program (C_AwareTTW), as well as age the sample member first became limited (C_DisAge and C_AdultChildOnset), ever worked for pay (C_EvrWorked), and worked

when limited (C_WrkdWhenLim). Similarly because G46 (family paid for services) was skipped if family was indicated as a source of payment in G45 (who paid for services from provider), constructs were created to identify sources of payments across these items (C_SelfFamPayServ-C_C_AgencyPayServ). The constructed variable code included in the codebooks provides the original questionnaire variables used to create each constructed variable.

Finally, several constructed variables were created in section G to summarize information across providers. In order to facilitate reporting of services received, respondents were asked to list the names of places where they received various types of services (employment, job training, medical services, mental health services, and schooling). For each provider mentioned, respondents were then asked whether they received services from this provider in 2004. To consolidate this information, constructs were created to flag whether each particular type of services was ever received (C_EvrUseEmploy, C_EvrUsedServ) and which specific services were received in 2005 across providers (C_PhyTh2005-C_JobCch2005). Additionally, constructs were created to flag whether services were ever received from particular types of providers (for example, C_EvrUseSVR) and whether those providers were used in 2005 (for example, C_UseSVR2005). The provider constructs created in section G are discussed in more detail below.

5. Scales

Constructed variables were also created to summarize items that were part of a pre-existing scale. This included creating a total SF-8^{TM} physical and mental score (C_PCS8TOT, C_MCS8TOT), a score on the CAGE alcohol scale (C_CAGEAlcohol), and a drug dependence indicator (C_DrugDep). A body mass index (C_BMI) construct was also created based on height and weight.

6. Other

Additional constructs were created to simplify analysis of income data (by creating a poverty level construct), impairments (by creating a series of variables to identify the number of ADL, IADL, physical, emotional, and other impairment types), job information (by collapsing information across jobs), and information about Employment Networks and length of time in the TTW program (by summarizing across Employment Networks).

E. SSA ADMINISTRATIVE DATA

MPR received administrative data from SSA for the purposes of selecting the sample; contacting, locating, and verifying sample members; and to fill information or drive instrument pathing in the survey instrument. Personally identifying information received from SSA (for example, Social Security number, name, address, and telephone number) is not included on either the Restricted Use or Public Use data file. Key items that were used for the creation of sampling strata and those that were used to dictate pathing in the instrument are included. These variables begin with –OrgSampInfo" to indicate that they are original sample file variables.

Because SSA benefit amount received last month was not asked of respondents, this information was retrieved from SSA administrative variables and was incorporated into the monthly income variables, C_AmtOthReg and C_TotGovCashBen. Additionally, back payments received from SSA were included as other income received on a nonregular basis for the variable C_AmtOthNonReg.

Additional administrative variables from SSA records were appended to the Public Use File to enable more comprehensive data analysis. These data retain their original names and are included at the end of the file.

F. PUBLIC USE VARIABLES

Some data were edited to ensure the confidentiality of survey respondents for the Public Use File. Editing for the Public Use File involved excluding variables containing information that could potentially be used to directly or indirectly identify a sample member, and constructing new variables to mask extreme or rare values and populations. Using the current OMB checklist on confidentiality as a guideline,[21] we developed encryption/masking algorithms that would maximize the analytic value of the data while maintaining acceptable confidentiality for the program participants. These variables were created for the Public Use file to mask identifying questionnaire data. These constructs end with a PUB and replace the original survey item on the Public Use file. These variables are also included on the Restricted Use File.

1. Variable Exclusion

In order to minimize the likelihood of indirect identification of a sample member, variables that could identify residents of smaller geographic areas or sample members possessing rare attributes (outliers) were deleted. Particular attention was paid to variables showing fewer than 100 sample members with a given characteristic (small cell sizes). The file was also simplified by dropping variables with little analytic value. These included survey administration variables, source variables that had corresponding imputed versions, imputation flags, source variables that were summarized in a constructed variable, and constructed variables that had not yet been utilized in round 1 or round 2 analyses. Data elements with quality problems that would reduce their analytic value were also dropped. SSA administrative data appended to the Restricted

[21] The Interagency Confidentiality and Data Access Group, working under the auspices of OMB's Federal Committee on Statistical Methodology, has developed a set of guidelines for statistical agencies to use in deciding whether statistical disclosure limitation procedures should be applied prior to releasing tabular and micro data. The latest version of the checklist is dated July 1999.

Access File were also dropped. In their place, select key administrative variables were masked and added to the file as new constructs. Appendix H provides a list of all variables dropped or replaced and the reason the variable was excluded. See Appendix A for a list of all variables included and dropped from the Public Use File.

2. Masking and the Construction of New Variables

The remaining variables were assessed for their confidentiality disclosure risk. When survey questions identified relatively rare populations, a new variable was constructed to combine small groups into larger groupings. For many variables that posed a potential risk, constructed variables summarizing the information already existed on the file. When constructed variables did not exist, MPR prepared masking algorithms that maximized their analytic value while maintaining acceptable confidentiality for the program participants. Masking algorithms included top and bottom coding of continuous variables, collapsing continuous variables into categories, and combining responses for categorical variables. These Public Use File constructs were assigned the same variable name as the source variable and end with a "PUB" to indicate that they were created for the Public Use File. A complete list of all variables edited for confidentiality with a brief description of the recode, can be found in Appendix I. Descriptions of the specific recodes and construct specifications for each variable can also be found in the codebook.

G. ADDITIONAL DETAILS ON SELECTED CONSTRUCTED VARIABLES

1. Jobs Held in 2005

In section C (Current Employment), job-related information was collected for each job held at the time of interview. In section D (Jobs/Other Jobs in 2005), information was collected for any other jobs held in 2005 not already reported in section C. Data for each job are represented

on the Restricted Use data file with an _n indicating which job the data are in reference to (for example, D6mth_1 indicating month started first job, D4mth_2 indicating month started second job, and so on). In both sections, respondents were asked to report first on their main job, that is, the job at which they worked the most hours, and then to subsequently report on other jobs held. To reduce respondent burden, respondents were not asked to report on any jobs held during 2005 that had previously been mentioned in section C as current employment. Rather, employment data from section C were copied to section D items during data processing for all current jobs also held during the 2005 time period. See Table V.8 for a list of all job-specific items that were filled in with section C data. Items in section D that had no equivalent in section C (D8mtn, D8yr, D23) were coded as .L (indicating logical skip).

TABLE V.8

JOB VARIABLES IN SECTIONS C AND D

Variable in C	Variable in D	Variable Description
C2	D4	Occupation
C3	D5	Industry
C4mth, C4yr	D6mth, D6yr	Start month and year of job
No equivalent item	D8mth, D8yr	Stop month and year of job
C6	D14	Self-employed status
C7	D15	Sheltered workshop status
C8	D16	Hours usually worked per week
C9	D17	Weeks usually worked per year
C10	D18	Paid by the hour
C11	D19	Hourly pay
C12amt, C12hop	D20amt, D20hop,	Amount of pre-tax pay
C13amt, C13hop	D21amt, D21hop	Amount of post-tax pay
No equivalent item	D23_1 thru D23_22	Reasons for stopping work

Source: NBS, round 3.

a. Including Current Jobs Held in 2004 in Section D

Jobs mentioned in section C were defined as held in 2005 if C4yr (year started current job) was earlier than or equal to 2005 and the job held in 2005 had been held for longer than one month. Each applicable job from section C was copied into the first blank job slot in section D (for example into D6mth_2 if D6mth_1 already contained data and into D6mth_3 if both D6mth_1 and D6mth_2 already contained data). The variables C_job_from_SecC_1 through C_job_from_SecC_4 are included on the Restricted Access data file to indicate which jobs from section C (by job number) were copied into specific section D job slots.

b. Determining the Main Job Held in 2005

In addition to copying job data from section C to the section D items, it was necessary to determine which job held in 2005 was the main job. Prior to including the jobs from section C, the main job held in 2005 was stored as job 1. Since it was possible that a job reported in section C was the respondent‘s main job in 2005, hours worked in 2005 on each job were compared with the first job mentioned in section D once the jobs from section C were incorporated. The job with the greatest number of hours per year (numbers of hours per week multiplied by the number of weeks per year), was considered the main 2005 job. [22] The variable Main_Job_grid_num identifies the job number of the main job held in 2005 after this analysis.

[22] If hours per year could not be calculated due to missing data on either number of hours per week or number of weeks per year, it was coded as missing. If hours per year was missing for all 2005 section C jobs, job 1 in section D was counted as the main job in 2005. If there were no jobs listed in section D, and hours per year was missing for all 2005 jobs in section C, the first job listed in C that was a 2005 job was counted as the main job in 2005. If hours per year was missing for job 1 in section D, the section C job with most hours per year was counted as the main 2005 job.

If there was no 2005 job from section C, or hours per year was missing for all section C 2005 jobs, job 1 in section D was counted as the main 2005 Job. If hours per year was missing for all 2005 section C jobs and job 1 in section D, job 1 in section D was counted as the main job in 2005.

This was used to create a series of variables ending with _m representing each job specific item listed in Table III.5 for the main job held in 2005 (for example D6mth_m and D6yr_m). It is important to note that information related to the main job was not deleted from the job_1-job_5 variables when this was done. For example, for a case in which three jobs are listed in section D (after copying relevant jobs from section C) and the second job is determined to be the main job, information related to hours worked on this job will be found in both C8_m and in C_8_2. Therefore, _m jobs should not be counted as additional jobs. On the public use version of the file, only the main job variables (_m) are provided for jobs held in 2005.

For the purposes of the constructed variables created in this section, separate constructs were created for each job mentioned (job 1, job 2, and so on). Additional constructs were created for the ~~m~~main" job (C_MainJob2005SOC, C_MainJob2005NAICS, C_MainJobHrPay2005, C_MainJobMnthPay2005, C_MainJobMnthPayTH2005, and C_MnthsMain2005Job) as identified by the variable Main_Job_grid_num. As stated above, information in the main job constructs is replicated in one of the other job slots on the restricted file and does not represent an additional job.

2. Service Providers

In section G, respondents were asked to discuss employment-related services and supports they received in 2005. To aid in the recall of employment-related services received in 2004, respondents were first asked if they had ever received employment services, job training, medical services, or counseling to improve their ability to work or live independently. For each type of service, respondents were asked to list up to eight providers or places where the service was received (at G2, G11, G16, and G20). Provider type was then collected for each provider mentioned. To minimize respondent burden by avoiding the need to ask provider type again if a provider was listed under two or more services, interviewers could indicate that a provider had

already been mentioned, thus skipping the provider type follow-up questions. Once a list of providers ever used was obtained, respondents were asked when they last received services from each provider. Follow-up questions regarding specific services received, number of visits, duration of visits, cost of services, and usefulness of services received in 2005 were asked for each provider from whom services were received in 2005.

Data for each specific provider mentioned were stored in a grid using the convention _n (1-34) to indicate which provider the data were associated with. Providers mentioned under G2 (employment services received) were stored in slots _1-_10; providers mentioned under G11 (job training) were stored in slots _11-_18; providers mentioned under G16 (medical services) were stored in slots _19-_26; and providers mentioned under G20 (therapy or counseling) were stored in slots _27-_34. This convention was maintained throughout the section so that data associated with the second provider listed under G2 (_2) are always found in the _2 variables (for example G33_2) and data associated with the second provider listed under G11 (_12) are found in the 12 variables (for example G33_12).

To simplify this section for the purposes of analyses, a series of constructed variables were created. To start, each provider was assigned a provider type code (C_ProvType2005_1-_34) indicating the type of provider services were received from (see the NBS codebook for detailed construct specifications). Constructs were also created to identify services received from each provider (for example C_PhyTh2005_1-_34, C_OccTh2005_1-_34, and so on), the duration of the visit with each provider (C_DurProvVisit_1- C_DurProvVisit_34), the number of contacts with each provider (C_NumProvCont_1- C_NumProvCont_34), and the usefulness of services received from each provider (C_ProvUse2005_1- C_ProvUse2005_34). Variables were then created to classify providers by type so that a list of providers and services received by provider type could be developed. For example, if the first provider mentioned was a state vocational

rehabilitation agency (SVRA), this provider was considered the first SVRA provider (C_Provtype2005_01_1) with _01 indicating provider type 1 (SVRA) and _1 indicating first provider of this type mentioned. If the second provider was a mental health provider, this provider was considered the first mental health provider (C_Provtype2005_03_1). If the third provider was another SVRA, this provider was considered to be the second SVRA provider (C_Provtype2005_01_2). These variables were then linked to data pertaining to specific services received, payment of services, and duration and usefulness of visits mentioned above (for example C_Phyth2005_01_1 indicating that physical therapy was received by the first SVRA provider). Provider types were classified as shown in Table V.9.

TABLE V.9

NUMERIC VALUES ASSOCIATED WITH PROVIDER TYPES

Provider Type	Description
1	SVRA
2	Welfare Agency
3	Mental Health Agency
4	Other State Agency
5	Private Business
6	Other Non-State Agency
7	School
8	Unemployment Office
9	Unknown Employment/Training
10	Clinic/Hospital/MD
11	Rehabilitation Treatment Center
12	Other Medical/Mental Health Provider
13	Unknown Medical/Mental Health Provider

Source: NBS, round 3.

Additional constructs were created that summarized provider types across services, for example C_EvrUseSVR, (indicating that the sample member ever used an SVRA), and

C_UseSVR2005, (indicating that an SVRA was used in 2005); and specific services received across providers, for example C_PHYTH2005 (received physical therapy in 2005), C_OCCTHER2005 (received occupational therapy in 2005), and so on.

For the Public Use File, the source variables and intermediary constructs related to the data collection grid (_1-_34) are not provided. Due to small cell sizes, welfare agency (type=2), other non-state agency (type=6), and unemployment office providers (type=8) were combined with provider type other state agency (type=4). Provider type=9 (unknown employment provider) and type=13 (unknown medical provider) were dropped. For the Public Use File, second and third providers for many provider types were dropped due to small cell sizes.

H. ITEMS SKIPPED FOR LONGITUDINAL RESPONDENTS

Several items in the round 3 survey were not asked of longitudinal respondents who had completed a prior round. These items were skipped because they were no longer relevant, because answers should be stable across time (for example race), or because the information obtained would overlap with previous responses. In most cases, these items were coded as .N (not applicable) if cases would otherwise be eligible to receive the item. Where answers should be stable over time, these items were logically imputed with respondents‘ answers at round 1 to simplify analyses (age and year first became limited, limited before age 18, worked for pay when limited, job when first limited required computer use, ethnicity, race, and education of parents). Table V.10 provides a summary of items that were not asked of longitudinal respondents who completed round 1 or round 2 and describes how each item was coded at round 3 for these cases. It should be noted that although the data file does not include a flag indicating which cases are coded using round 1 or round 2 data, analysts can use the variable R3_statusR1R2R3 (response status-round 1, round 2, and round 3) to determine which cases are coded with round 1 or round 2 data for longitudinal participants on a given item (R3_STATUSR1R2R3=2, 3, or 4).

Constructed variables such as C_disage, C_evrworked, and C_adultChild_Onset incorporate these edits. Appendix C provides the rationale for why each item was not asked for longitudinal cases.

TABLE V.10

ITEMS SKIPPED FOR LONGITUDINAL RESPONDENTS WHO COMPLETED ROUND 1 OR 2

Variable	Variable Label	Description of Coding
B9	Received benefits in last 5 years	Coded as .N if R3_statusR1R2R3=2, 3 or 4 and B5=0
B11	Still have conditions that made eligible	Coded as .N if R3_statusR1R2R3 =2, 3 or 4 and B5=0 and B9=1
B13	Previously eligible for other reasons	Coded as .N if R3_statusR1R2R3 =2, 3 or 4 and B5=0 and B9=1 and B11=1
B18_age	Age first became limited	If R3_statusR1R2R3 =2 or 3, populated with round 2 data; if 4, populated with R1 data
B18_year	Year first became limited	If R3_statusR1R2R3=2 or 3, populated with round 2 data; if 4, populated with R1 data
B19	Limited before 18	If R3_statusR1R2R3=2 or 3, populated with round 2 data; if 4, populated with R1 data
B22	Working for pay when first limited	If R3_statusR1R2R3=2 or 3, populated with round 2 data; if 4, populated with R1 data
B23	Job when first limited required computer use	If R3_statusR1R2R3=2 or 3, populated with round 2 data; if 4, populated with R1 data
E3	Ever heard of PASS	Coded as .N if (bstatus=1 or bstatus=3) and R3_statusR1R2R3=2, 3, or 4 and FIXE2=00 and E3 = (.) and E5 = (.) and E7 = (.) and E9 in (.).
E5	Ever heard of earned income exclusion	Coded as .N if (bstatus=1 or bstatus=3) and R3_statusR1R2R3=2, 3, or 4 and FIXE2=00 and E3 = (.) and E5 = (.) and E7 = (.) and E9 in (.).
E7	Ever heard of PESS	Coded as .N if (bstatus=1 or bstatus=3) and R3_statusR1R2R3=2, 3, or 4 and FIXE2=00 and E3 = (.) and E5 = (.) and E7 = (.) and E9 = (.).
E9	Ever heard of Continued Medicaid Eligibility	Coded as .N if (bstatus=1 or bstatus=3) and R3_statusR1R2R3=2, 3, or 4 and FIXE2=00 and E3 = (.) and E5 = (.) and E7 = (.) and E9 = (.).
E12	Ever heard of student earned-income exclusion	Coded as .N if (Orgsampinfo_bstatus=1 or 3) and orgsampinfo_age is <=25 and SSIAGE<=22 and R3_statusR1R2R3=2, 3, or 4 and FIXE2=00 and E12 = (.) .

TABLE V.10 *(continued)*

Variable	Variable Label	Description of Coding
E15, E17	Ever heard of Trial Work Period	Coded as .N if (bstatus=2 or bstatus=3) and R3_statusR1R2R3=2, 3, or 4 and FIXE14=00 and E15 = (.) and E17 = (.).
E17	Ever heard of Extended Period of Eligibility for Medicare	Coded as .N if (bstatus=2 or bstatus=3) and R3_statusR1R2R3=2, 3, or 4 and FIXE14=00 and E15 = (.) and E17 = (.) .
E19, E20a, E20c	Ever heard of Impairment-Related Work Expenses	Coded as .N if R3_statusR1R2R3=2, 3, or 4 and E19 = (.) and E20a = (.) and E20c = (.) .
E20a	Ever heard of Expedited Reinstatement	Coded as .N if R3_statusR1R2R3=2, 3, or 4 and E19 = (.) and E20a = (.) and E20c = (.).
E20c	Ever heard of BPAOs	Coded as .N if R3_statusR1R2R3=2, 3, or 4 and E19 = (.) and E20a = (.) and E20c = (.) .
E26	Year heard about TTW	Coded as .N if R3_statusR1R2R3=2, 3, or 4 and (E21=1 or E24=1 or E25=1).
E27	Received info in the mail about TTW	Coded as .N if R3_statusR1R2R3=2, 3, or 4 and (E21=1 or E24=1 or E25=1).
E28_1-_10	Who sent TTW info	Coded as .N if R3_statusR1R2R3=2, 3, or 4 and (E21=1 or E24=1 or E25=1).
E29	Someone called about TTW	Coded as .N if R3_statusR1R2R3=2, 3, or 4and (E21=1 or E24=1 or E25=1).
E30_1-_10	Who called about TTW	Coded as .N if R3_statusR1R2R3=2, 3, or 4and (E21=1 or E24=1 or E25=1).
E31	Someone talked to about TTW	Coded as .N if R3_statusR1R2R3=2, 3, or 4and (E21=1 or E24=1 or E25=1).
E32_1-_10	Who talked to about TTW	Coded as .N if R3_statusR1R2R3=2, 3, or 4and (E21=1 or E24=1 or E25=1).
E33	Learned about TTW on website	Coded as .N if R3_statusR1R2R3=2, 3, or 4and (E21=1 or E24=1 or E25=1).
E34	Got Ticket in mail	Coded as .N if R3_statusR1R2R3=2, 3, or 4and (E21=1 or E24=1 or E25=1).
E35	Tried to get Ticket	Coded as .N if R3_statusR1R2R3=2, 3, or 4and (E21=1 or E24=1 or E25=1).
E36	Ever used ticket	Coded as .N if R3_statusR1R2R3=2, 3, or 4and (E21=1 or E24=1 or E25=1) and (E34=1 or E35=1).
E48	Ever used Ticket with any other EN	Coded as .N if R3_statusR1R2R3=2, 3, or 4 and E37a ne 1 and E41_1-E41_4 ne 1 and E21=1 or E24=1 or E25=1.
E49	Number ENs ever signed up with	Coded as .N if R3_statusR1R2R3=2, 3, or 4, and E37a ne 1 and E41_1-E41_4 ne 1 and E21=1 or E24=1 or E25=1.

TABLE V.10 *(continued)*

Variable	Variable Label	Description of Coding
E50mth	Month first used Ticket with other EN	Coded as .N if R3_statusR1R2R3=2, 3, or 4and E37a ne 1 and E41_1-E41_4 ne 1 and E21=1 or E24=1 or E25=1.
E50yr	Year first used Ticket with other EN	Coded as .N if R3_statusR1R2R3=2, 3,or 4 and E37a ne 1 and E41_1-E41_4 ne 1 and E21=1 or E24=1 or E25=1.
L1	Ethnicity	If R3_statusR1R2R3=2, 3 or 4, then field populated with R1 data. If R1_final=19 or R1_final=29, then field populated with .D.
L2	Race	If R3_statusR1R2R3=2, 3 or 4, then field populated with R1 data. If R1_final=19 or R1_final=29, then field populated with .D.
L4	Highest grade mother completed	If R3_statusR1R2R3=2, 3, or 4 then field populated with R1 data. If R1_final=19 or R1_final=29, then field populated with .D.
L5	Highest grade father completed	If R3_statusR1R2R3=2, 3, or 4 then field populated with R1 data. If R1_final=19 or R1_final=29, then field populated with .D.

Source: NBS, round 3.

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

VI. SAMPLING WEIGHTS

The final analysis weights for the Representative Beneficiary Sample and the Ticket Participant Sample[23] were determined via a four-step process: (1) calculate the initial 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. This chapter describes these computations for both the Representative Beneficiary Sample and the Ticket Participant Sample. Section A summarizes the procedures used to compute and adjust the sampling weights, and the procedure for creating composite weights. (Composite weights were used in all rounds to combine the Representative Beneficiary Sample and the Ticket Participant Cross-Sectional Sample, and to combine two samples in the Ticket Participant Sample.) Procedures for computing the weights for the Representative Beneficiary Sample are described in detail in Section B. Sections C and D cover the same information for the Ticket Participant Cross-Sectional Sample and the Ticket Participant Longitudinal Sample, respectively.

A. COMPUTING AND ADJUSTING THE SAMPLING WEIGHTS: A SUMMARY

1. Representative Beneficiary Sample

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

[23] The "Ticket Participant Sample" in this chapter refers to both the Ticket Participant Cross-Sectional and Ticket Participant Longitudinal Samples.

beneficiaries.[24] We used four age-based strata in each PSU. In particular, beneficiaries were stratified into the following age groups: 18- to 29-year-olds, 30- to 39-year-olds, 40- to 49-year-olds, and 50- to 64–year-olds. Because we used a composite size measure to select the PSUs, we could achieve equal probability samples in the age strata and nearly equal workload in each PSU for the Representative Beneficiary Sample.[25]

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

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

[25] 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).

the actual sample size. The release-adjusted weights were post-stratified to population totals obtained from SSA.[26]

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

We calculated this adjustment factor in two stages: (1) estimating a propensity score for locating a sample member, and (2) estimating a propensity score for response among located sample members. In our experience with this survey, factors associated with the inability to locate a person tend to be different from factors associated with cooperation. The unlocated

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

person cannot deliberately avoid or otherwise refuse to cooperate. For instance, that person may have chosen not to list his or her number, or may frequently move from one address to another, but he or she has not shown a specific unwillingness to cooperate with the survey itself. Located nonrespondents may deliberately avoid the interviewer, or may be expressing displeasure or hostility toward surveys in general, or SSA in particular.

To develop the logistic propensity models for round 3, we used information from the SSA data files and geographic information (such as urban/rural or region) as covariates. Using a liberal level of statistical significance (0.3) in forward and backward stepwise regression models, we made an initial attempt to reduce the pool of covariates and interactions. We used a higher significance level because the purpose of the model was to improve the estimation of the propensity score, not to identify statistically significant factors related to response. In addition, the information sometimes reflected proxy variables for some underlying variable that was both unknown and unmeasured. Any covariate or interaction that clearly was unrelated to locating the respondent, or to response propensity, was excluded from the pool.

The next step was to carefully evaluate a series of models by comparing the following measures of predictive ability and goodness of fit: the R-squared statistic,[27] Akaike's Information Criterion (AIC),[28] percentage of concordant and discordant pairs,[29] and the Hosmer-

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

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

[29] A pair of observations is concordant if a responding subject has a higher predicted value than the nonresponding subject, discordant if not, and tied if both members of the pair are either respondents, nonrespondents, or have the same predicted values. It is desirable to have as many concordant and as few discordant pairs as is possible (Agresti 1990).

Lemeshow goodness-of-fit test.[30] Model-fitting also involved reviewing the statistical significance of the coefficients of the covariates in the model and avoiding any unusually large adjustment factors. In addition, we also avoided data warnings in SUDAAN.[31] We then used the specific covariate values for each located person (cooperating person) to estimate a propensity to be located (to cooperate), from which we calculated the adjusted weights. The location-adjusted weight is the product of the released adjusted weight and the inverse of the location propensity score; the nonresponse-adjusted weight is the product of the location-adjusted weight and the inverse of the cooperation propensity score.

Once the adjustments were made, we trimmed the survey weights (if necessary) to avoid unusually large weights, which would make the survey estimates less precise. We used the design effect attributed to the variation in the sampling weights as a statistical measure to determine both the necessity and the 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, the design effect due to unequal weighting was checked within the age-related sampling strata, and trimming to reduce the design effect was employed in only one age stratum (the 30- to 39–year-old age group). The design effect was reduced from

[30] 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 can be used to test goodness of fit even when some of the covariates are continuous (Hosmer and Lemeshow 1989).

[31] 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 avoiding the first two was of the highest priority. These warnings almost always were caused by a response cell with a count that was too small, which required dropping covariates or collapsing categories in covariates.

1.09 to 1.06, which was the maximum design effect among all the age strata in the Representative Beneficiary Sample.

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[32]). After post-stratification, we checked the survey weights again to determine whether more trimming was necessary. In round 3, trimming was not necessary after post-stratification in the Representative Beneficiary Sample.

2. Ticket Participant Cross-Sectional Sample

The initial sampling cross-sectional weights for the Ticket Participant Cross-Sectional Sample were computed 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 the sample members and, when possible, selected an initial larger (augmented) sample. For participants in Phase 2 states using either the milestone-outcome or the outcome-only payment system, and for participants in Phase 3 states using the outcome-only payment system, the PSUs in the initial sampling design did not have enough participants to support analysis tasks—even with all participants in the PSUs from these two payment types selected for the sample. As a result, it was necessary to supplement the sample from the PSUs with a second independent sample of Ticket participants from two geographic strata defined by the PSUs (participants residing in a PSU, or not residing in any of the PSUs). The sample members within the initial sample design are referred to as the clustered sample; members of the second independent sample are referred to as the unclustered sample. Sample members in the unclustered sample

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

were randomly selected from the entire population of milestone-outcome and outcome-only participants in Phase 2 states, and from the entire population of outcome-only participants in Phase 3 states, in the two aforementioned geographic strata.[33] The combination of data from the clustered and unclustered samples to calculate estimates is referred to as a "paired sample design," and is discussed later in this document.

As with the Representative Beneficiary Sample, we computed the weights for the augmented sample and then adjusted them for the number of sample members that were in the final sample.[34] We adjusted for nonresponse separately for located sample members, and then for response among these sample members. Because Ticket participants were generally easier to locate due to their participation in the Ticket program, the number of Ticket participants who could not be located was very small. Hence, for all Ticket participants except those in Phase 2 states using the traditional payment system, we calculated the location adjustment using the weighting class method. However, the location adjustment for Phase 2 Ticket participants using the traditional payment system, and the response adjustments for located Ticket participants of both phases and all three payment types, were calculated using logistic propensity models. The modeling procedures were similar to those used with the Representative Beneficiary Sample.

The size of the sample for the three payment types was similar, but the size of the population for each was very different. (More than 80 percent of the population of Ticket participants used the traditional payment system. Specific percentages for each phase and payment type are given

[33] Because of the small populations for the payment types where the paired sample design was required, Ticket participants who resided in the selected PSUs for these payment types often were selected for both the clustered and the in-PSU strata of the unclustered samples. Hence, these duplicate cases had to be accounted for in the weighting process, as discussed later.

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

in Section C.) Hence, the sampling weights differed substantially in magnitude from one payment system to the next. As a result, we conducted the weight adjustments separately for each payment type. For the subsamples associated with each phase and payment type within the Ticket Participant Cross-Sectional Sample, we trimmed the weights to ensure that the design effect due to unequal weighting was not substantially greater than 3.0 (less than 3.0, if possible). (More details about the trimming of the participants‘ weights, and the design effects due to unequal weighting before and after trimming, are given in Section C.) The final adjustment for the participants‘ weights was a post-stratification adjustment to the counts of participants within subgroups defined by age and gender in the sampling frame. After post-stratification, we checked the survey again to determine whether more trimming was necessary. In round 3, although trimming was required before post-stratification in the Ticket Participant Cross-Sectional Sample, no trimming was required after post-stratification.

3. Composite Cross-Sectional Weights

Although the Ticket participant population constitutes a small subset of the beneficiary population, some analyses require a sample with a substantial number of individuals both within and outside the Ticket participant population. This can be accomplished by combining the Ticket Participant Cross-Sectional Sample and Representative Beneficiary Sample and using composite weights to account for the fact that the samples have been combined. When conducting analyses representing the beneficiary population, these weights can be used to make estimates about participants within the beneficiary population. (Analyses limited to the participants subpopulation use weights from the Ticket Participant Cross-Sectional Sample only.)

In round 1, we used a sophisticated procedure to create these weights, such that the variance of survey estimates was minimized. This procedure allowed for weights to be applied to observations that were duplicated across the two samples.[35] However, because the Ticket participants were such a small fraction of the beneficiary sample frame, we used a simpler alternative method in rounds 2 and 3.

In round 3, composite weights were developed only for the Phase 2 and Phase 3 Ticket participants in the Representative Beneficiary Sample. As indicated earlier, the round 3 Ticket Participant Sample included only Phase 1 Ticket participants who were selected in round 1. This meant that no cross-sectional sample of all Phase 1 participants at round 3 was available. Of the 55 Ticket participants in the Representative Beneficiary Sample, only 33 were Phase 2 or Phase 3 cases. Of these 33 cases, 31 were respondents (that is, they had completed interviews, or were ineligible after sample selection in round 3). These respondents included 14 from Phase 2 states and 17 from Phase 3 states. We replaced the original Representative Beneficiary Sample weights with a value of zero among these 31 cases. To ensure that the Ticket participant population would be represented, we replaced these members of the Representative Beneficiary Sample with the 3,161 members of the Ticket Participant Cross-Sectional Sample (2,091 from Phase 2 states and 1,070 from Phase 3 states) with completed interviews (or ineligible dispositions after sample selection).[36] From the Ticket participants sampling frame, there were 33,500 participants from Phase 2 states and 31,023 participants from Phase 3 states. Because the sum of the weights of the 31 Ticket participants from Phase 2 or Phase 3 states in the Representative Beneficiary

[35] A complex procedure also was used to combine the clustered and unclustered samples of the Ticket Participant Sample in all rounds. This procedure is described in Section C of this chapter.

[36] This does not include sample members who were selected for the rounds 1 or 2 Ticket Participant Samples, were no longer Ticket participants in round 3, but were sampled in round 3 anyway for longitudinal purposes.

Sample did not equal the sampling frame totals for each phase, we computed post-stratification adjustments for the remaining beneficiary weights. The sum of the weights for the 31 participants in the Representative Beneficiary Sample is an unbiased estimate of the number of participants in the sampling frame. Because of the relatively small sample size, this estimate did not equal the known total in the sampling frame, as was expected. The post-stratification adjustment realigned the population totals.

4. Ticket Participant Longitudinal Sample

For longitudinal analyses, the inferential population is defined by the population at the time the Ticket program was rolled out for the group of states in question, and not by the Ticket participant population as it was constituted in round 3. For Phase 1 longitudinal cases, the inferential population is the set of Phase 1 Ticket participants at round 1. For Phase 2 longitudinal cases, the inferential population is the set of Phase 2 Ticket participants at round 2. We conducted a nonresponse bias analysis to evaluate the differences between Phase 1 cases who responded in various combinations of rounds (round 1, round 2, and/or round 3). We concluded that no systematic differences in selected key variables were apparent between groups of Phase 1 cases differentiated by their response patterns. Based on this result, and on consultations with SSA and TTW Project Part A contract staff,[37] we determined that three sets of Ticket participant longitudinal weights were sufficient for anticipated longitudinal analyses. These longitudinal weights included: a set for Phase 1 participants who responded[38] in rounds 1,

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

[38] "Respondents" include individuals who had completed interviews or were ineligible after sample selection (i.e., when surveyed).

2, and 3; a set for Phase 1 participants who responded in rounds 1 and 2, and a set for Phase 2 participants who responded in rounds 2 and 3. The initial sampling longitudinal weights for the Ticket Participant Longitudinal Sample were computed on the basis of the inverse of the selection probability for the participants in the first round for Phase 1 longitudinal weights and the second round for Phase 2 longitudinal weights. For the calculation of longitudinal weights, Ticket participants from Phase 2 states using the outcome-only and milestone-and-outcome payment systems needed a paired sampling design. Similarly, a paired sampling design was also required for the calculation of longitudinal weights for Ticket participants from Phase 1 states using the outcome-only payment system.

As with the cross-sectional weights, we calculated adjustments for nonresponse in two stages: (1) a location adjustment for locating a sample member, and (2) a cooperation adjustment for response among located sample members. However, unlike the cross-sectional weight nonresponse adjustments, we used logistic models for the location and cooperation adjustments for all payment types and phases. The inverse of the propensity score was used as the adjustment factor. The adjusted weight for each sample case is the product of the initial sampling weight and the adjustment factor.

We trimmed the weights so that the design effect due to unequal weighting was not substantially greater than 3.0 (less than 3.0, if possible), and post-stratified them to add up to the round 1 frame totals for Phase 1 cases, and round 2 frame totals for Phase 2 cases. (Details about the trimming used with longitudinal weights, and the design effects before and after trimming, are given in Section C.)

5. 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 and longitudinal samples, as well as the composite weights. For the sake of objectivity, we chose a statistician who was not directly involved in the project.

B. REPRESENTATIVE BENEFICIARY SAMPLE

1. Initial Weights

The initial weights were computed using the inverse of the probability of selection. For the Representative Beneficiary Sample, samples were selected independently in each of four age strata in each geographic unit or PSU.[39] The number of sample members selected in each stratum and PSU for the augmented sample was determined by allocating three times the target sample size across the 84 geographic units (PSUs and secondary sampling units) independently for each stratum.[40] This ensured that plenty of reserve sample units were available in case response or eligibility rates were lower than expected. The augmented sample size for the three younger age strata (18 to 29 years, 30 to 39 years, and 40 to 49 years) was 4,000 sample members (roughly three times the target sample size of 1,333); for beneficiaries 50 to 64 years, the augmented sample size was 3,000 (again, three times the target sample size of 1,000). By using the composite size measure described previously, the initial weights for the full augmented sample of 15,000 sample members were calculated by taking the inverse of the global sampling rate (F_i) for each stratum. The global sampling rates and initial weights are given in Table VI.1.

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

[40] An augmented sample that was three times as large as needed was selected to allow for 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.

As described previously, the full sample was randomly partitioned into subsamples called waves that mirrored the characteristics of the full sample. The waves were formed in each of the four sampling strata in the 84 geographic units (a total of 336 combinations of PSU and sampling strata). At the start of data collection, a preliminary sample was assigned to the data collection effort, and additional waves were assigned as needed, based on experience with eligibility and response rates. Within the 336 combinations of PSU and sampling strata, the initial weights were adjusted to account for the number of waves assigned to data collection. The final sample size for the Representative Beneficiary Sample was 3,382 beneficiaries, as shown under "Released Sample" in Table VI.1.

TABLE VI.1

SURVEY POPULATION AS OF JUNE 30, 2005, INITIAL AUGMENTED SAMPLE SIZES AND INITIAL WEIGHTS BY SAMPLING STRATA IN THE NATIONAL BENEFICIARY SURVEY

Sampling Strata (ages as of June 30, 2005)	Survey Population[a]	Augmented Sample Size	Global Sampling Rate (Fj)	Initial Sample Weights	Released Sample
Beneficiaries between 18 and 29 years old	1,064,845	4,000	0.003756	266.2	943
Beneficiaries between 30 and 39 years old	1,271,121	4,000	0.003147	317.8	941
Beneficiaries between 40 and 49 years old	2,514,758	4,000	0.001591	628.7	935
Beneficiaries between 50 and 64 years old	5,534,098	3,000	0.000542	1844.7	563
Total	**10,384,822**	**15,000**			**3,382**

Source: Sample allocation and counts computed by MPR.

[a]The survey population represents all SSI and SSDI beneficiaries.

2. Nonresponse Adjustment

As in virtually all surveys, the sampling weights must be adjusted to compensate for sample members that cannot be located or who, once located, refuse to respond. First, weighted logistic regression models were fitted where the binary response was whether the sample member could be located. Using variables obtained from SSA databases, a pool of covariates from which to

choose a final location model was selected through stepwise regression. This pool included both main effects and interactions. From this pool of covariates, candidate models were compared using various measures of goodness of fit and predictive ability, while avoiding large adjustments. This process was repeated for interview respondents among the located sample members, where another weighted logistic regression model was fitted. The two levels in the binary response for this model were ―respondent" 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 determined to be ineligible after sample selection (an ineligible respondent). Ineligible sample members included persons who were never SSA beneficiaries, were in the military service at the time of the survey, were incarcerated, had moved outside of the United States, or were deceased at the time of the survey.

Using the procedures outlined above, the main factors or attributes affecting our ability to locate and interview the sample member included the personal characteristics of the sample member (race, ethnicity, gender, and age), the type of beneficiary (recipient of SSI, SSDI, or both), identity of the payee with respect to the beneficiary, whether the beneficiary and the applicant for benefits lived in the same location, primary disability classification, type of disability claim (a person with a disability, a survivor, or other), living situation of beneficiary, and geographic characteristics.

a. Coding of Survey Dispositions

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

1. Eligible respondents.
2. Ineligible respondents (sample members who were ineligible after sample selection, including deceased, sample members in the military or incarcerated, sample members living outside of the United States, and other ineligibles).
3. Located nonrespondents (including active or passive refusals and language barrier situations).
4. Unlocated sample members (sample members who could not be located using 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.[41]

b. Response Rates

The 81.1 percent response rate for the Representative Beneficiary Sample quoted in the introduction to this document is the **weighted overall completion rate**, given in the first line of Table VI.2. This response rate is the weighted count of sample members for whom a completed interview was obtained or who were determined to be ineligible, divided by the weighted sample count of all sample members. [42] It can be determined by taking the product of the weighted

[41] Disposition codes 420 (institutionalized) and 430 (unavailable during field period) were classified as nonrespondent codes in rounds 2 and 3, even though they were considered ineligible codes in round 1. This affected 8 cases in the round 2 beneficiary sample and 6 cases in the round 3 beneficiary sample. As a result, the nonresponse adjusted weight for these cases was 0 in rounds 2 and 3, 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 very small.

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

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

TABLE VI.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	3,382	3,186	95.2	2,723	85.2	81.1
SSI Only, SSDI Only, or Both SSI and SSDI						
SSI only	1,507	1,393	92.8	1,195	84.4	78.2
SSDI only	1,181	1,137	97.2	947	84.0	81.7
Both SSI and SSDI	694	656	94.1	581	90.9	85.4
SSI or SSDI						
SSI only, or in both SSI & SSDI programs	2,201	2,049	93.2	1,776	86.4	80.5
SSDI only, or in both SSI & SSDI programs	1,875	1,793	96.4	1,528	85.7	82.6
Constructed Disability Status						
Deaf	39	37	97.8	28	75.6	73.7
Mental	1,822	1,706	94.3	1,442	83.0	78.3
Physical	1,360	1,296	96.5	1,124	86.7	83.6
Unknown	161	147	87.1	129	90.6	78.6
Beneficiary's Age (Four Categories)						
18-29 years	943	874	92.7	775	88.6	82.2
30-39 years	941	878	93.3	728	82.7	77.4
40-49 years	935	894	95.6	760	85.2	81.3
50-64 years	563	540	95.9	460	85.2	81.7
Sex						
Male	1,762	1,650	94.2	1,405	85.3	80.4
Female	1,620	1,536	96.2	1,318	85.2	81.9
Hispanicity						
Non-Hispanic	3,215	3,039	95.5	2,588	84.8	81.0
Hispanic	167	147	88.1	135	95.1	83.9
Race						
White	1,989	1,895	96.3	1,630	85.6	82.4
Black	781	725	93.8	621	86.7	81.2
Unknown	547	505	92.8	431	84.8	78.8
Asian American, Pacific Islander	44	42	97.8	24	40.9	39.2
North American Indian or Alaskan Native	21	19	92.4	17	93.9	87.1
Living Situation						
Living alone	1,941	1,812	93.5	1,571	86.3	80.6
Living with others	1,40	129	94.1	105	79.4	75.1
Living with parents	34	31	91.3	28	90.5	82.6
In institution or unknown	1,267	1,214	96.8	1,019	84.5	81.8

TABLE VI.2 *(continued)*

	Sample	Located Sample		Response among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Did the Applicant for Benefits Live in Same Zip Code as Beneficiary?						
No	303	265	86.8	224	85.3	73.8
Yes	1,596	1,507	95.0	1,319	88.2	83.8
No information	1,483	1,414	96.4	1,180	83.0	79.9
Identity of the Payee with Respect to the Beneficiary						
Beneficiary received beneficiary payments himself or herself	1,964	1,846	95.2	1,567	84.7	80.6
Payee is a family member	1,106	1,043	95.6	902	87.1	83.3
Payee is an institution	220	210	94.3	183	86.8	82.1
Other	92	87	93.2	71	84.3	78.2
Changes in Telephone Number						
No changes in last 5 years	58	55	95.6	47	88.2	84.0
One change in last 5 years	5	5	100.0	4	70.8	70.8
Two or more changes in last 5 years	2	2	100.0	2	100.0	100.0
No information on phone number	3,317	3,124	95.2	2,670	85.2	81.1
Number of Moves in Last 5 Years						
No moves in last 5 years	39	37	96.6	34	92.6	89.3
One or more moves in last 5 years	2	2	100.0	2	100.0	100.0
No information on number of moves	3,341	3,147	95.2	2,687	85.2	81.1
Type of Claim						
Survivor	310	299	94.4	261	88.5	83.4
Disabled	1,610	1,533	96.6	1,300	85.3	82.4
Unknown	1,462	1,354	92.9	1,162	84.4	78.3
Census Region						
Midwest	801	763	95.6	674	88.6	84.7
Northeast	550	517	94.6	417	81.7	77.2
South	1,370	1,289	94.6	1,117	86.1	81.4
West	661	617	96.4	515	82.4	79.5
Census Division						
East North Central	594	565	94.9	502	89.1	84.5
East South Central	306	294	97.3	249	86.3	83.8
Middle Atlantic	382	358	94.2	296	84.7	79.7
Mountain	175	163	96.5	148	89.1	86.0
New England	168	159	95.7	121	74.9	71.5
Pacific	486	454	96.4	367	80.0	77.1
South Atlantic	729	680	93.4	590	84.8	79.2
West North Central	207	198	97.6	172	87.1	85.3
West South Central	335	315	94.8	278	89.1	84.4

TABLE VI.2 *(continued)*

	Sample	Located Sample		Response among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Metropolitan						
Metropolitan areas of 1 million population or more	1,474	1,378	94.5	1,148	84.3	79.5
Metropolitan areas of 250,000 to 999,999 population	864	809	95.1	695	83.6	79.5
Metropolitan areas of less than 250,000 population	381	362	93.3	316	88.1	82.1
Nonmetropolitan areas adjacent to large metropolitan areas	242	233	98.4	209	90.7	89.2
Nonmetropolitan areas adjacent to medium or small metropolitan areas	252	244	98.3	214	88.6	87.1
Nonmetropolitan areas not adjacent to metropolitan areas	169	160	97.3	141	82.8	80.5

Source: NBS, round 3.

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 VI.2 as 95.2 percent. The **weighted cooperation rate** (the weighted completion rate among located sample members), 85.2 percent in Table VI.2, is the weighted count of sample members for whom a completed interview was obtained, or who were determined to be ineligible, divided by the weighted sample count of all located sample members. Weighted cooperation rates reflect the common survey situation that once a person is located, repeated contact efforts often will result in a completed interview.

The weighted rates are used because (1) the sampling rates (therefore the sampling weights) vary substantially across the sampling strata, as seen in Table VI.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 to use either in the data analysis or to determine as ineligible for the analysis.

c. Factors Related to Location and Response

In addition to overall response rate information, Table VI.2 also provides information for selected factors associated with locating a sample member, and factors associated with response among located sample members. The table includes the unweighted counts of all sample members, counts of located sample members, and counts of sample members for whom a completed interview was obtained, or who were determined to be ineligible. The table also includes the weighted location rate, the weighted completion rate among the located sample members, and the weighted overall completion rate for these factors, which helped to inform the decision about the final set of variables used in the nonresponse adjustment models.

d. Propensity Models for Weight Adjustments

The response propensity models used to determine the nonresponse adjustments were developed using the main effects described previously, plus selected interactions. To identify candidate interactions among these variables for the modeling, we first ran a chi-squared automatic interaction detector (CHAID) analysis in SPSS to find possible significant interactions. CHAID normally is attributed to Kass (1980) and Biggs et al. (1991), and its application in SPSS is described in Magidson (1993). The CHAID procedure iteratively segments a data set into mutually exclusive subgroups that share similar characteristics based on their effect on nominal or ordinal dependent variables. It automatically checks all variables in the data set and creates a hierarchy that shows all statistically significant subgroups. The algorithm finds splits in the population, which are as different as possible based on a chi-square statistic. It is a forward stepwise procedure; it finds the most diverse subgrouping, and then each of these subgroups is split further into more diverse sub-subgroups. Sample size limitations are set to avoid generating cells with small counts. It stops when splits no longer are significant; that is, that group is homogeneous with respect to variables not yet used, or when the cells contain too

few cases. The CHAID procedure results in a tree that identifies the set of variables and interactions among the variables that have an association with the ability to locate a sample member (and the propensity of a located sample member to either respond or be ineligible). CHAID first was run with all covariates, then rerun a few times with the top variable in the tree removed, to ensure all potentially important interactions were retained for further consideration. The resulting pool of covariates was reduced further 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 either were located or unlocated (for the location models), complete or not complete (for the cooperation models), or the total number of sample members at that level was fewer than 20, then levels were collapsed if collapsing was possible. If collapsing was not possible, then the covariate or interaction was excluded from the pool.[43]

All of the resulting candidate main effects, and the interactions identified using CHAID, were then processed using forward and backward stepwise regression (using SAS Logistic procedure with weights normalized to the sample size) to further refine the candidate variables and interaction terms.[44] After identifying a smaller pool of main effects and interactions for potential inclusion in the final model, a set of models was evaluated carefully to determine the final model. Because the SAS logistic procedure does not incorporate the sampling design, the final selection of the covariates was accomplished using the logistic regression procedure in SUDAAN.

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

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

For selecting variables or interactions in the stepwise procedures, we included variables or interactions that had a statistical significance level (alpha level) of 0.30 or lower (instead of the commonly used 0.05).[45] Once the candidate list of main effects and interactions was determined, a thorough model-fitting process was used to determine a parsimonious model with few very small propensities. Model selection criteria were described in the overview of this chapter (Section A). The variables used in the model as main effects and interactions are summarized in Table VI.3 for locating a sample member and in Table VI.4 for cooperation among located sample members. The R-squared is 0.054 (0.168 when rescaled to have a maximum of 1) for the location model and 0.078 (0.138 when rescaled) for the cooperation model.[46] These values are similar to those observed for other response propensity modeling efforts using logistic regression with design-based sampling weights. For the location model, the percentage of concordant pairs is 69.1 percent, 29.7 percent of the pairs are discordant[47], and the p-value for the chi-square statistic from the Hosmer-Lemeshow (H-L) goodness-of-fit test is 0.184[48]; these values indicate a reasonably good fit of the model to the data. For the cooperation model, the percentage of concordant pairs is 65.7 percent, and 33.7 percent of pairs are discordant. The p-value for the

[45] As stated earlier, we used a higher significance level because the purpose of the model was to improve the estimation of the propensity score, and not to identify statistically significant factors related to response. In addition, the information sometimes reflected proxy variables for some underlying variable that was both unknown and unmeasured.

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

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

[48] 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 can be used to test goodness of fit even when some of the covariates are continuous (Hosmer and Lemeshow 1989).

chi-square statistic for the (H-L) goodness-of-fit test is 0.420 for this model. Since the Akaike's Information Criterion (AIC) is a relative number, and has no meaning on its own, values for the AIC are not provided here. [49]

TABLE VI.3

LOCATION LOGISTIC PROPENSITY MODEL: REPRESENTATIVE BENEFICIARY SAMPLE

Factors in the Location Model
Main Effects
DIG_1
REPREPAYEE_1
GENDER (SEX)
METRO_1
DIVISION_1
SSI_SSDI
RACE_1
TOC_1
AGECAT
PDZIPSAME_1
Two-Factor Interactions
DIVISION_1*RACE_1
DIG_1*PDZIPSAME_1
DIG_1*TOC_1
DIVISION_1*DIG_1
SSI_SSDI*DIVISION_1
SSI_SSDI*REPREPAYEE_1
DIVISION_1*SEX
AGECAT*TOC_1
AGECAT*DIG_1
AGECAT*DIVISION_1

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

TABLE VI.4

COOPERATION LOGISTIC PROPENSITY MODEL: REPRESENTATIVE BENEFICIARY SAMPLE

Factors in the Cooperation Model
Main Effects
AGECAT_2
RACE_2
HISPANICITY
METRO_2
DIVISION_2
SSI_SSDI_
GENDER (SEX)
REPREPAYEE_2
PDZIPSAME_2
DIG_2
TOC_2
LIVING_2
Two-Factor Interactions
DIG_2*RACE_2
DIVISION_2* RACE_2
SSI_SSDI * RACE_2
DIVISION_2* PDZIPSAME_2
DIG_2* PDZIPSAME_2
METRO_2* TOC_2
DIVISION_2*METRO_2
DIG_2*DIVISION_2
TOC_2* DIVISION_2
SSI_SSDI* DIVISION_2
SSI_SSDI* METRO_2
METRO_2*SEX
DIVISION_2*SEX
PDZIPSAME_2*METRO_2
REPREPAYEE_2*METRO_2
DIG_2*SSI_SSDI
SSI_SSDI*SEX

The primary factors are identified by the base variable, often followed by the suffix ―1.‖ If the levels associated with the variable as it is used in the location model correspond directly to those in Table VI.2, no suffix is given. However, if levels of the variable used in the location model are collapsed from those shown in Table VI.2, the base variable name is followed by the suffix ―1.‖

The factors with levels used in the location model include:

1. ***DIG_1.*** Disability diagnostic classification; three levels: (1) mental disability, (2) physical disability (including deaf cases), and (3) unknown.
2. ***REPREPAYEE_1.*** The identity of the payee with respect to the beneficiary; two levels: (1) the beneficiary received benefit payments from a family member, and (2) an institution received payments on behalf of the beneficiary, or the beneficiary received benefit payments himself or herself, or identity of payee not known.
3. ***GENDER (SEX).*** Two levels: (1) Male, and (2) Female.
4. ***METRO_1.*** Urbanicity of beneficiary's place of residence; four levels: (1) beneficiary lived in metropolitan area, (2) beneficiary lived in nonmetropolitan area adjacent to a metropolitan area of 1 million or more, and (3) beneficiary lived in nonmetropolitan area adjacent to a metropolitan area of less than 1 million, (4) beneficiary lived in nonmetropolitan area not adjacent to metropolitan area.
5. ***DIVISION_1.*** Geographic region (based on U.S. Census divisions) of beneficiary's place of residence; four levels: (1) South Atlantic, (2) West (Mountain and Pacific), (3) Midwest (East North Central and West South Central), (4) all other divisions.
6. ***SSI_SSDI.*** Beneficiary status; three levels: (1) SSI only, (2) SSDI only, (3) Both SSI and SSDI.
7. ***RACE_1.*** Race; two levels: (1) White, (2) Not white or not known to be white.
8. ***TOC_1.*** Type of claim; 2 levels: (1) Disability claim, (2) Survivor claim or unknown.
9. ***AGECAT.*** Beneficiary's age category; four levels: (1) age in range 18 to 29 years, (2) age in range 30 to 39 years, (3) age in range 40 to 49 years, and (4) age in range 50 to 64 years.
10. ***PDZIPSAME_1.*** 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, or information unknown.

Various interactions among these variables were also included in the model for locating the sample member. The main effects using the variable names listed above, as well as interactions, are provided in Table VI.3. An expanded form of Table VI.3, showing the specific levels of the interactions shown in Table VI.3, along with parameter estimates and their standard errors, is provided in Appendix J.

For the cooperation models, the primary factors are identified by the base variable, often followed by the suffix ―2." As with the location model, if the levels associated with the variable

used in the cooperation model are collapsed from those given in Table VI.2, the base variable name is accompanied by the suffix, but if no collapsing was necessary, no suffix is given. The factors include[50]:

1. ***AGECAT_2:*** Beneficiary's age category; 3 levels: (1) age in range 18 to 29 years, (2) age in range 30 to 39 years, and (3) age in range 40 to 64 years
2. ***RACE_2.*** Race of the beneficiary; four levels: (1) white, (2) black, (3) Asian or Pacific Islander, and (4) race known to be neither white nor black nor Asian/Pacific Islander, or unknown.
3. ***HISPANICITY.*** Whether the beneficiary was Hispanic or not; two levels: (1) Hispanic, and (2) not Hispanic, or unknown.
4. ***METRO_2.*** Urbanicity of beneficiary's place of residence; four levels: (1) beneficiary lived in metropolitan area with population of 1 million or more, (2) beneficiary lived in metropolitan area with population between 250,000 and 1 million, (3) beneficiary lived in metropolitan area with population less than 250,000, (4) beneficiary lived in nonmetropolitan area.
5. ***DIVISION_2.*** Geographic region (based on U.S. Census divisions) of beneficiary's place of residence; seven levels: (1) New England, (2) Middle Atlantic, (3) South Atlantic, (4) West South Central, (5) Mountain, (6) Pacific and (7) Midwest (East North Central and West North Central), and (8) East South Central.
6. ***SSI_SSDI.*** Beneficiary status; three levels: (1) SSI only, (2) SSDI only, (3) Both SSI and SSDI.
7. ***GENDER (SEX).*** Two levels: (1) Male, and (2) Female.

1. ***REPREPAYEE_2.*** The identity of the payee with respect to the beneficiary; two levels: (1) the beneficiary received benefit payments from a family member, and (2) an institution received payments on behalf of the beneficiary, or the beneficiary received benefit payments himself or herself, or identity of payee not known.
2. ***PDZIPSAME_2.*** Whether the beneficiary and the applicant for benefits lived in the same zip code; two levels: (1) beneficiary and applicant lived in the same zip code, (2) beneficiary and applicant lived in different zip codes/information unknown.
8. ***DIG_2.*** Disability diagnostic classification; three levels: (1) mental disability, (2) physical disability (excluding deaf cases), and (3) deafness/unknown.

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

9. ***TOC_2.*** Beneficiary's type of claim; two levels: (1) disability claim, and (2) survivor claim, or unknown.
10. ***LIVING_2.*** Beneficiary's living situation; three levels: (1) beneficiary lives alone, (2) beneficiary lives with others and (3) others/unknown.

Once again, various interactions among these variables were also included in the model for the cooperation of the sample members. The main effects using these variable names, as well as interactions, are provided in Table VI.4. An expanded form of Table VI.4, with the specific levels of the interactions shown in Table VI.4, along with parameter estimates and their standard errors, is provided in Appendix J.

After adjustments were applied to the sampling weights, the distribution of weights was reviewed to determine if trimming of the sampling weights was necessary. Prior to trimming, the maximum design effect due to unequal weighting was 1.09, observed with the second youngest age group stratum. Trimming reduced this design effect to 1.06, which was still the maximum design effect, due to unequal weighting among all the strata.

3. Post-Stratification

Post-stratification is the procedure in which the weighted sums of the response-adjusted weights are aligned to known totals external to the survey. This process offers face-validity for reporting population counts and has some statistical benefits. For the Representative Beneficiary Sample, we post-stratified to the 24 population totals obtained from SSA.[51] 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). No trimming was conducted after post-stratification.

[51] These totals were obtained from a frame file provided by the SSA, giving information on basic demographics for all SSI and SSDI beneficiaries.

C. TICKET PARTICIPANT CROSS-SECTIONAL SAMPLE

As noted earlier, the Ticket Participant Cross-Sectional Sample was selected from the round 3 population of Ticket-to-Work participants in Phase 2 and Phase 3 states, a subset of all SSI/SSDI beneficiaries, which was partitioned based on different payment types in the Ticket-To-Work payment system (traditional vocational rehabilitation, milestone-outcomes, and outcome-only). Ticket participants using the traditional payment system accounted for 84 percent (28,170 of 33,500) of Phase 2 participants and 84 percent (25,913 of 31,023) of Phase 3 participants at the time the sampling frame was developed. Participants using the milestone-outcomes payment system totaled 4,138 Phase 2 participants (12 percent of all Phase 2 participants) and 4,410 Phase 3 participants (14 percent of all Phase 2 participants). Phase 2 participants using the outcome-only payment system totaled only 1,192 Phase 2 participants (4 percent of all Phase 2 participants) and 700 Phase 3 participants (2 percent of all Phase 3 participants). As was also noted earlier, the PSUs in the initial sampling design did not contain a sufficient number of participants in the milestone-outcome payment type in Phase 2 states, or the outcome-only payment type in Phase 2 or Phase 3 states, to support analysis tasks. As a result, the clustered sample, consisting of respondents selected within the initial sample design, was supplemented by a sample randomly selected from the entire population of milestone-outcome and outcome-only participants in Phase 2 states, and from the entire population of outcome-only participants in Phase 3 states (the unclustered sample).

The clustered sample was part of the original sample design, so all of the respondents in the clustered sample were selected from within PSUs, whereas the unclustered sample included units that may or may not have been in the selected PSUs. The unclustered sample was therefore organized into two strata: in the PSU or not in the PSU. In most cases, the respondents who were selected for the in-PSU stratum of the unclustered sample were also in the clustered sample. The

weights for these duplicate cases had to be appropriately adjusted to account for a single respondent's appearance in two independent samples. The compositing scheme used to do this is discussed in the next subsection. In addition, respondents who could not be located by the central office[52] based on sample frame information were treated differently in the clustered and unclustered samples. In the clustered sample, potential respondents who could not be located were sent to the field for further follow-up so that personal interviews could be attempted. In the unclustered sample, no further attempt was made to locate potential respondents who could not be located by the central office. If a sample member was selected as part of both the clustered and unclustered samples, and was sent to the field for further follow-up and located in the field, the response had to be treated differently between the two samples. For the sample respondent, the value in the clustered sample was recorded according to its final status in the field, whereas the value in the unclustered sample was recorded as "ineligible for field follow-up." Sample members with no field follow-up (in the unclustered sample) were not "selected" for field follow-up. This process is analogous to the accepted practice of subsampling of nonrespondents for more intensive effort—in this case, we subsampled cases in the clustered sample for field follow-up. Ineligible-for-field-follow-up cases in the unclustered sample were treated differently than other ineligible cases, regardless of whether the observation was duplicated with a clustered observation. The procedure used to create composite weights (described in the next subsection) was not applied to these cases. Rather, such a case in the unclustered sample would have its weight zeroed out. If such a case was duplicated with one in the clustered sample, the clustered sample case kept its original weight, appropriately adjusted so that the sum of weights was kept the same. The final sample sizes for the participants cross-sectional sample are in Table VI.5.

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

Although a portion of the Ticket participant population in round 3 included Phase 1 cases, no cross-sectional sample of these Phase 1 cases was selected in round 3, as indicated by the N/A entries in Table VI.5.

TABLE VI.5

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

Sampling Strata (Payment System)	Survey Population[a]	Initial Augmented Sample Size[b]	Released Sample
Total Phase 1	44,265	N/A	N/A
1. Traditional payment type	39,357	N/A	N/A
2. Milestone-outcome payment type	3,613	N/A	N/A
3. Outcome-only payment type	1,295	N/A	N/A
Total Phase 2	33,500	3,388	2,797
1. Traditional payment type	28,170	1,000	867
2. Milestone-outcome payment type		1,500	937
Clustered sample	4,138	389	389
Unclustered sample	4,138	1,111	548
In PSUs	403	109	59
Not in PSUs	3,735	1,002	489
3. Outcome-only payment type		888	993
Clustered sample	1,192	123	123
Unclustered sample	1,192	765	870
In PSUs	123	79	88
Not in PSUs	1,069	686	782
Total Phase 3	31,023	2,858	1,373
1. Traditional payment type	25,913	1,000	444
2. Milestone-outcome payment type	4,410	1,000	444
3. Outcome-only payment type		858	485
Clustered sample	700	237	237
Unclustered sample	700	621	248
In PSUs	237	210	84
Not in PSUs	463	411	164

Source: Sample allocation and counts computed by MPR.

[a] This column reflects weighted totals before compositing.

[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 Phase 2 participants using the milestone-outcome or outcome-only payment types, and for Phase 3 participants using the outcome-only payment type, a paired sample design was employed, whereby the participants who were in the PSUs potentially could be selected for both samples.

For the clustered samples for TTW participants, the sample was allocated across the 79 PSUs, with the Los Angeles PSU receiving a double allocation because it had two selections. Because of the smaller population sizes, we used only the full PSUs; we did not use the 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

The initial weights were computed based on the probability of selection within the PSU of the augmented sample and the probability of selection for the PSU. For the unclustered sample for the milestone-outcome and outcome-only participants in Phase 2 states, and for the outcome-only participants in Phase 3 states, we computed the initial weights based on the selection probability within the two sampling strata (in one of the PSUs, or not in any PSU). Since only a portion of the augmented sample was actually released for use, the initial weights then were adjusted for the sample actually used in the survey.

2. Dual Frame Estimation

To obtain estimates for the outcome-only Ticket Participant Samples in Phase 2 and Phase 3 states, and to obtain estimates for the milestone-outcome Ticket Participant Sample in Phase 2 states, it was necessary to combine the clustered and unclustered samples using a ―paired sample design.‖ As noted earlier, if a potential respondent in the unclustered sample could not be located by the central office, he or she was considered ―ineligible for field follow-up‖ and no further attempts were made on that case. However, if a potential respondent was in the clustered sample and could not be located by the central office, the case was sent to the field for additional locating efforts (field follow-up). The paired sample design is the methodology used to combine

the samples while accounting for these different rules of field follow-up. This requires the creation of composite weights that can be applied to the combined samples.

a. Conceptual Framework for Composite Weights

To compute a survey estimate, *Est(Y)*, using information from both samples (such as the proportion who are currently working), one cannot simply combine the two samples without adjusting the weights, since the clustered and unclustered samples in the Ticket Participant Sample represent the same target population among the Ticket Participants. Separate estimates can be computed from each sample, within each payment type, and combined, using the equation

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

where *Y(clustered)* is the survey estimate from the clustered sample for the given payment type, *Y(unclustered)* is the survey estimate from the unclustered sample for the given payment type, and λ is an arbitrary constant between 0 and 1. For example, for the Phase 2 milestone-outcomes payment type in the round 3 data, there were 389 in the clustered sample and 548 in the unclustered sample. The estimates to be combined are the proportion of the 389 in the clustered sample who are currently working and the proportion of the 548 in the unclustered sample who are currently working. In practice, of course, it is more complicated than this, because we have to account for the different rules used in the two samples for following up with nonrespondents or unlocated sample members, as will be discussed later. For the sampling variance, *V(Y)*, the estimate is computed using the equation

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

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

value of λ will result in an unbiased estimate of the survey estimate, but not necessarily an estimate with the minimum sampling variance. A lambda value producing a sampling variance at its minimum value results in the shortest confidence interval and, by implication, the most precise point estimate.

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

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

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

In this case, the minimum variance is:

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

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

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

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

where *n(clustered)* and *n(unclustered)* are the sample sizes of the clustered and unclustered central office-located samples respectively, and *deff(clustered)* and *deff(unclustered)* are the design effects due to unequal weighting for the clustered and unclustered central office-located samples, respectively.

b. Application of Composite Weights to Ticket Participant Sample

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

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

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

For units in the unclustered central office-located sample:

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

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

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

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

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

3. Nonresponse Adjustment

As with the Representative Beneficiary Survey, the sampling weights were adjusted in two stages, one stage for the sample members who could not be located and another stage for those who, once located, refused to respond. Due to the small number of unlocated Ticket participants, the location adjustment was calculated using the weighting class method for all except traditional Phase 2 participants. However, the location adjustment for traditional Phase 2 Ticket participants and the response adjustment for located Ticket participants of both phases and all three payment types was calculated using logistic propensity models. For the milestone-outcome and outcome-only payment types in Phase 2 states, and for the outcome-only payment type in Phase 3 states, the nonresponse adjustments were applied to the composite weights for the clustered and

unclustered samples. Roughly equal sample sizes with vastly different population sizes for the three payment types resulted in substantial differences in the magnitude of the weights. Thus, it was necessary to calculate separate adjustments for each payment type and phase, first for the location adjustment and subsequently for the cooperation adjustment. This resulted in a total of 12 weight adjustments, including 5 location adjustments using the weighting class method, and 7 adjustments using logistic propensity models. The models were fitted in the same way as the adjustment models for the Representative Beneficiary Sample, as described in Section B.2 of this chapter. The main factors or attributes affecting our ability to locate and interview Ticket Participant sample members were the same as those used to locate and interview Representative Beneficiaries, where the specific covariates for each of the 12 weight adjustments varied as described in subsequent sections.

a. Coding of Survey Dispositions

The scheme used to code respondents included the four general categories described in Section B.2: eligible respondents, ineligible respondents, located nonrespondents, and unlocated sample members.[53]

b. Response Rates

The response rate for the Ticket Participant Cross-Sectional Sample is 84.4 percent, which is the weighted overall completion rate for Phase 2 and Phase 3 cases. This rate is a combination of the Phase 2 weighted overall completion rate (84.5 percent) and the Phase 3 weighted overall

[53] Disposition codes 420 (institutionalized) and 430 (unavailable during field period) were classified as nonrespondent codes in round 3, even though they were considered ineligible codes in round 1. This affected 4 cases in the round 3 participant sample. As a result, the nonresponse adjusted weight for these 4 cases was 0 in round 3, 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 very small.

completion rate (84.2 percent). It is also the product of the weighted location rate and the weighted completion rate among located sample members. The weighted location rate is 97.0 percent, the combination of the Phase 2 and Phase 3 location rates (97.1 percent and 96.8 percent, respectively). The weighted cooperation rate (the weighted completion rate among located sample members), is 87.0 percent, the combination of the Phase 2 and Phase 3 weighted completion rates (87.1 percent and 86.9 percent, respectively).

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

c. Factors Related to Location and Response

Tables VI.6 - VI.11 provide information for selected factors associated with locating a sample member within each phase-payment type combination, and factors associated with response among located sample members. The tables include unweighted counts of all sample members, counts of located sample members, and counts of the sample members for whom a completed interview was obtained, or who were determined to be ineligible. The tables also include the weighted location rate, the weighted completion rate among located sample members, and the weighted overall completion rate for these factors, which helped inform the decision about the final set of variables used to define the weighting classes and in the nonresponse adjustment models.

TABLE VI.6

WEIGHTED LOCATION AND RESPONSE RATES FOR TICKET PARTICIPANT CROSS-SECTIONAL SAMPLE, PHASE 2 STATES, MILESTONE-OUTCOME PAYMENT SYSTEM, BY SELECTED CHARACTERISTICS

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
All	841	824	98.2	674	81.2	79.8
SSI Only, SSDI Only, or Both SSI and SSDI						
SSI only	245	234	95.2	191	78.9	75.2
SSDI only	375	369	98.7	306	81.9	80.9
Both SSI and SSDI	221	221	100.0	177	82.1	82.0
SSI or SSDI						
SSI only, or in both SSI & SSDI programs	466	455	97.7	368	80.6	78.8
SSDI only, or in both SSI & SSDI programs	596	590	99.2	483	82.0	81.3
Constructed Disability Status						
Deaf	32	31	99.1	17	58.7	58.6
Mental	498	482	96.8	390	80.4	77.9
Physical	297	297	100.0	254	84.4	84.4
Unknown	14	14	100.0	13	90.8	90.6
Beneficiary's Age (Four Categories)						
18-29 years	228	224	98.4	173	76.8	75.5
30-39 years	189	180	95.5	148	83.1	79.4
40-49 years	227	223	98.7	188	83.7	82.7
50-64 years	197	197	100.0	165	82.1	82.0
Sex						
Male	422	414	98.7	333	81.7	80.7
Female	419	410	97.7	341	80.8	78.9
Hispanicity						
Hispanic	11	10	88.0	9	86.3	76.3
Non-Hispanic/unknown	830	814	98.3	665	81.2	79.8
Race						
White	431	426	98.7	348	81.5	80.5
Black	323	312	97.3	261	82.5	80.1
Other/unknown	87	86	98.5	65	75.6	74.5
Living Situation						
Living alone	457	446	97.7	366	83.1	81.2
Living with others/unknown	384	378	98.7	308	79.2	78.2
Did the Applicant for Benefits Live in Same Zip Code as Beneficiary?						
No	98	95	97.2	81	82.2	80.1
Yes	472	460	98.0	371	81.9	80.3
No information	271	269	98.8	222	79.8	78.9

TABLE VI.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	508	496	98.2	424	85.6	84.0
Payee is a family member	261	258	98.5	197	75.8	74.8
Payee is an institution	54	53	98.2	41	73.7	72.4
Other	18	17	93.6	12	60.0	56.9
Changes in Telephone Number						
No changes in last 5 years	249	245	97.9	205	87.4	85.4
One or more changes in last 5 years	13	12	90.8	10	80.2	72.8
No information/other	579	567	98.6	459	78.0	77.0
Number of Moves in Last 5 Years						
No moves in last 5 years	123	120	97.3	106	90.0	87.4
One or more moves in last 5 years	10	10	100.0	7	60.0	59.6
No information on number of moves	708	694	98.4	561	79.8	78.6
Type of Claim						
Survivor	90	90	100.0	71	78.6	78.7
Disabled	483	476	98.8	393	82.8	81.8
Unknown	268	258	96.3	210	78.9	76.0
Census Region						
Midwest	381	372	98.4	307	81.7	80.3
Northeast	33	33	100.0	24	71.1	71.2
South	413	405	97.7	333	82.1	80.3
West	14	14	100.0	10	71.4	71.4
Census Division						
East North Central	338	331	99.2	276	82.1	81.4
East South Central	76	75	99.0	65	85.5	84.6
Middle Atlantic	16	16	100.0	13	82.4	82.2
Mountain	12	12	100.0	8	66.6	66.7
New England	17	17	100.0	11	60.4	60.5
Pacific	2	2	100.0	2	100.0	100.0
South Atlantic	47	47	100.0	40	84.9	84.9
West North Central	43	41	94.0	31	79.4	74.1
West South Central	290	283	96.5	228	79.4	76.8
Metropolitan						
Metropolitan areas of 1 million population or more	523	506	96.0	415	80.3	77.3
Metropolitan areas of 250,000 to 999,999 population	89	89	100.0	73	83.2	83.3
Metropolitan areas of less than 250,000 population	101	101	100.0	84	80.8	80.9
Nonmetropolitan areas adjacent to large metropolitan areas	56	56	100.0	49	88.0	87.8
Nonmetropolitan areas adjacent to medium or small metropolitan areas	42	42	100.0	29	68.8	69.0
Nonmetropolitan areas not adjacent to metropolitan areas	30	30	100.0	24	80.4	80.3

TABLE VI.6 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Longitudinal Sample Case						
Yes	372	362	97.2	293	82.2	79.8
No	469	462	99.1	381	80.4	79.7

Source: NBS, round 3.

TABLE VI.7

WEIGHTED LOCATION AND RESPONSE RATES FOR TICKET PARTICIPANT CROSS-SECTIONAL SAMPLE, PHASE 2 STATES, OUTCOME-ONLY PAYMENT SYSTEM, BY SELECTED CHARACTERISTICS

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
All	870	857	95.8	675	77.2	74.1
SSI Only, SSDI Only, or Both SSI and SSDI						
SSI only	111	108	97.3	81	75.9	73.4
SSDI only	665	658	98.2	516	76.0	75.0
Both SSI and SSDI	94	91	80.4	78	86.6	69.6
SSI or SSDI						
SSI only, or in both SSI & SSDI programs	205	199	88.4	159	81.0	71.4
SSDI only, or in both SSI & SSDI programs	759	749	95.6	594	77.3	74.2
Constructed Disability Status						
Deaf	21	21	100.0	16	79.8	78.0
Mental	457	449	98.0	360	75.7	74.8
Physical	385	380	92.6	294	79.1	73.1
Unknown	7	7	100.0	5	72.2	71.3
Beneficiary's Age (Four Categories)						
18-29 years	89	87	98.1	73	83.9	82.3
30-39 years	195	192	98.5	140	71.1	70.5
40-49 years	328	323	92.3	249	80.3	73.8
50-64 years	258	255	97.6	213	75.4	74.8
Sex						
Male	436	433	99.3	326	75.6	75.1
Female	434	424	92.5	349	78.5	73.2
Hispanicity						
Hispanic	15	14	92.8	13	92.3	85.5
Non-Hispanic/unknown	855	843	95.8	662	76.9	73.9
Race						
White	594	588	99.2	460	77.2	76.8
Black	173	170	94.4	131	72.9	69.1
Unknown	103	99	78.8	84	84.2	66.6
Living Situation						
Living alone	229	223	89.0	181	81.2	72.3
Living with others/unknown	641	634	98.1	494	75.9	74.8
Did the Applicant for Benefits Live in Same Zip Code as Beneficiary?						
No	65	65	100.0	58	90.4	90.5
Yes	330	323	90.4	255	78.6	71.0
No information	475	469	99.0	362	74.7	74.3

TABLE VI.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	716	704	95.0	555	77.2	73.6
Payee is a family member	105	104	99.1	79	74.3	73.9
Payee is an institution	29	29	100.0	28	96.5	96.4
Other	20	20	100.0	13	65.2	65.2
Changes in Telephone Number						
No changes in last 5 years	339	334	93.6	245	70.5	66.4
One or more changes in last 5 years	5	5	100.0	4	80.6	80.0
No information/other	526	518	97.3	426	82.2	79.9
Number of Moves in Last 5 Years						
No moves in last 5 years	91	90	98.9	70	78.1	77.2
One or more moves in last 5 years	5	5	100.0	4	80.2	80.0
No information on number of moves	774	762	95.4	601	77.0	73.8
Type of Claim						
Survivor	48	46	96.5	39	83.2	80.5
Disabled	722	714	95.6	559	76.4	73.3
Unknown	100	97	97.0	77	80.5	77.6
Census Region						
Midwest	160	156	88.0	127	72.7	64.8
Northeast	499	493	98.9	385	78.0	77.2
South	183	180	96.4	141	79.6	76.7
West	28	28	100.0	22	78.9	78.5
Census Division						
East North Central	71	68	94.1	56	56.9	55.7
East South Central	77	76	98.8	57	78.5	76.9
Middle Atlantic	92	90	97.7	69	76.3	74.5
Mountain	24	24	100.0	19	79.2	79.1
New England	407	403	99.1	316	78.4	77.8
Pacific	4	4	100.0	3	77.1	75.0
South Atlantic	61	60	98.4	48	78.7	77.9
West North Central	89	88	83.7	71	85.5	71.1
West South Central	45	44	89.3	36	83.2	74.5
Metropolitan						
Metropolitan areas of 1 million population or more	320	315	96.3	252	80.2	77.2
Metropolitan areas of 250,000 to 999,999 population	355	353	99.6	274	77.6	77.4
Metropolitan areas of less than 250,000 population	48	48	100.0	37	81.0	80.9
Nonmetropolitan areas adjacent to large metropolitan areas	19	17	64.8	13	91.5	58.6
Nonmetropolitan areas adjacent to medium or small metropolitan areas	82	79	96.4	62	78.8	75.6
Nonmetropolitan areas not adjacent to metropolitan areas	46	45	98.6	37	49.8	52.1

TABLE VI.7 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Longitudinal Sample Case						
Yes	412	405	94.1	300	71.6	67.7
No	458	452	97.4	375	82.8	80.6

Source: NBS, round 3

TABLE VI.8

WEIGHTED LOCATION AND RESPONSE RATES FOR TICKET PARTICIPANT CROSS-SECTIONAL SAMPLE, PHASE 2 STATES, 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	867	840	97.0	742	88.4	85.7
SSI Only, SSDI Only, or Both SSI and SSDI						
SSI only	234	226	96.5	199	87.4	84.3
SSDI only	416	407	98.0	359	88.5	86.6
Both SSI and SSDI	217	207	95.7	184	89.1	85.2
SSI or SSDI						
SSI only, or in both SSI & SSDI programs	451	433	96.1	383	88.2	84.7
SSDI only, or in both SSI & SSDI programs	633	614	97.2	543	88.7	86.2
Constructed Disability Status						
Deaf	35	31	88.9	24	77.3	68.7
Mental	486	471	96.8	418	88.7	85.9
Physical	332	324	97.9	290	89.8	87.9
Unknown	14	14	100.0	10	68.9	69.0
Beneficiary's Age (Four Categories)						
18-29 years	246	240	97.5	214	89.2	87.0
30-39 years	182	175	96.7	151	85.4	82.5
40-49 years	249	240	96.2	208	87.0	83.6
50-64 years	190	185	97.6	169	91.7	89.6
Sex						
Male	454	442	97.4	387	87.7	85.5
Female	413	398	96.5	355	89.0	85.9
Hispanicity						
Hispanic	10	10	100.0	8	80.6	80.6
Non-Hispanic/unknown	857	830	97.0	734	88.4	85.7
Race						
White	552	536	97.1	472	88.3	85.7
Black	201	196	97.7	175	88.5	86.4
Other/unknown	114	108	95.4	95	88.2	84.2
Living Situation						
Living alone	444	426	96.0	381	89.3	85.7
Living with others/unknown	423	414	98.0	361	87.4	85.6
Did the Applicant for Benefits Live in Same Zip Code as Beneficiary?						
No	78	71	91.2	61	85.3	77.5
Yes	473	462	97.6	412	89.0	86.9
No information	316	307	97.4	269	88.1	85.8

TABLE VI.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	534	518	97.0	457	88.3	85.7
Payee is a family member	261	252	96.7	227	90.0	86.9
Payee is an institution	51	51	100.0	47	92.9	92.8
Other	21	19	93.1	11	60.1	55.7
Changes in Telephone Number						
No changes in last 5 years	311	301	97.0	262	87.2	84.6
One or more changes in last 5 years	8	8	100.0	8	100.0	100.0
No information/other	548	531	96.9	472	88.8	86.1
Number of Moves in Last 5 Years						
No moves in last 5 years	122	118	97.0	103	87.9	85.2
One or more moves in last 5 years	7	7	100.0	7	100.0	100.0
No information/Other	738	715	97.0	632	88.3	85.6
Type of Claim						
Survivor	90	89	99.2	81	91.2	90.4
Disabled	553	534	96.8	471	88.5	85.7
Unknown	224	217	96.5	190	86.8	83.8
Census Region						
Midwest	449	435	96.8	383	88.5	85.6
Northeast	46	46	100.0	38	82.9	83.2
South	324	311	96.4	276	88.0	84.7
West	48	48	100.0	45	93.8	93.9
Census Division						
East North Central	368	358	97.3	314	87.9	85.5
East South Central	82	76	93.6	68	87.0	81.4
Middle Atlantic	12	12	100.0	10	84.2	84.2
Mountain	48	48	100.0	45	93.8	93.9
New England	34	34	100.0	28	82.6	82.9
Pacific	0	0	N/A	0	0	N/A
South Atlantic	158	156	98.7	134	86.0	84.8
West North Central	81	77	95.0	69	90.6	86.1
West South Central	84	79	94.1	74	93.7	88.2
Metropolitan						
Metropolitan areas of 1 million population or more	386	377	97.8	323	86.0	84.1
Metropolitan areas of 250,000 to 999,999 population	123	120	97.5	107	88.8	86.7
Metropolitan areas of less than 250,000 population	118	112	95.8	106	94.5	90.5
Nonmetropolitan areas adjacent to large metropolitan areas	119	113	94.8	100	88.3	83.7
Nonmetropolitan areas adjacent to medium or small metropolitan areas	26	25	95.3	22	85.9	82.0
Nonmetropolitan areas not adjacent to metropolitan areas	95	93	97.9	84	90.3	88.4

TABLE VI.8 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Longitudinal Sample Case						
Yes	432	420	97.5	367	87.3	85.0
No	435	420	96.5	375	89.4	86.3

Source: NBS, round 3.

TABLE VI.9

WEIGHTED LOCATION AND RESPONSE RATES FOR TICKET PARTICIPANT CROSS-SECTIONAL SAMPLE, PHASE 3 STATES, MILESTONE-OUTCOME PAYMENT SYSTEM, BY SELECTED CHARACTERISTICS

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
All	444	426	95.9	369	87.8	84.3
SSI Only, SSDI Only, or Both SSI and SSDI						
SSI only	157	151	96.5	132	88.7	85.6
SSDI only	169	162	95.1	141	88.0	83.8
Both SSI and SSDI	118	113	96.2	96	86.4	83.2
SSI or SSDI						
SSI only, or in both SSI & SSDI programs	275	264	96.4	228	87.7	84.5
SSDI only, or in both SSI & SSDI programs	287	275	95.6	237	87.3	83.6
Constructed Disability Status						
Deaf	8	8	100.0	4	62.6	63.2
Mental	246	232	95.0	196	86.4	82.1
Physical	182	178	96.8	162	91.3	88.4
Unknown	8	8	100.0	7	88.1	88.1
Beneficiary's Age (Four Categories)						
18-29 years	85	81	95.9	72	90.6	86.9
30-39 years	104	100	96.5	80	81.8	79.0
40-49 years	145	138	94.3	121	89.0	84.1
50-64 years	110	107	97.5	96	89.7	87.5
Sex						
Male	223	213	95.0	182	86.2	81.9
Female	221	213	96.8	187	89.3	86.5
Hispanicity						
Hispanic	14	14	100.0	13	93.1	92.9
Non-Hispanic/unknown	430	412	95.8	356	87.6	84.0
Race						
White	179	171	95.5	150	89.9	85.9
Black	186	178	95.7	152	85.6	81.9
Other/unknown	79	77	97.4	67	87.1	85.0
Living Situation						
Living alone	264	254	96.5	221	88.2	85.1
Living with others/unknown	180	172	95.0	148	87.2	83.0
Did the Applicant for Benefits Live in Same Zip Code as Beneficiary?						
No	42	37	88.5	30	81.4	72.0
Yes	225	220	98.0	194	89.7	87.9
No information	177	169	94.7	145	86.7	82.2

TABLE VI.9 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Identity of the Payee with Respect to the Beneficiary						
Beneficiary received beneficiary payments himself or herself	318	306	96.0	267	88.5	85.0
Payee is a family member	95	93	98.2	75	82.4	80.9
Payee is an institution	27	24	88.5	24	100.0	88.5
Other	4	3	75.6	3	100.0	75.6
Type of Claim						
Survivor	23	22	96.1	19	87.1	84.0
Disabled	269	258	95.6	224	87.8	84.0
Unknown	152	146	96.5	126	87.9	84.8
Census Region						
Midwest	50	47	94.6	42	88.6	83.8
Northeast	55	55	100.0	47	89.2	89.3
South	193	184	94.9	170	92.9	88.2
West	146	140	95.9	110	79.0	75.8
Census Division						
East North Central	33	30	90.9	28	93.4	84.9
East South Central	40	38	95.0	37	97.3	92.5
Middle Atlantic	55	55	100.0	47	89.2	89.3
Mountain	2	2	100.0	1	50.6	50.0
New England	0	0	N/A	0	0	N/A
Pacific	144	138	95.8	109	79.3	76.0
South Atlantic	119	114	94.8	105	92.8	88.0
West North Central	17	17	100.0	14	82.4	82.4
West South Central	34	32	94.9	28	88.9	84.3
Metropolitan						
Metropolitan areas of 1 million population or more	242	233	96.3	201	86.2	83.1
Metropolitan areas of 250,000 to 999,999 population	188	180	96.0	155	86.5	83.0
Metropolitan areas of less than 250,000 population	5	5	100.0	5	100.0	100.0
Nonmetropolitan areas adjacent to large metropolitan areas	2	1	56.1	1	100.0	56.1
Nonmetropolitan areas adjacent to medium or small metropolitan areas	7	7	100.0	7	100.0	100.0
Nonmetropolitan areas not adjacent to metropolitan areas	0	0	N/A	0	0	N/A

Source: NBS, round 3.

TABLE VI.10

WEIGHTED LOCATION AND RESPONSE RATES FOR TICKET PARTICIPANT CROSS-SECTIONAL SAMPLE, PHASE 3 STATES, OUTCOME-ONLY PAYMENT SYSTEM, BY SELECTED CHARACTERISTICS

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
All	427	409	94.9	326	78.2	74.0
SSI Only, SSDI Only, or Both SSI and SSDI						
SSI only	111	104	93.0	87	82.6	76.7
SSDI only	221	217	97.4	171	76.7	74.4
Both SSI and SSDI	95	88	90.0	68	77.7	70.1
SSI or SSDI						
SSI only, or in both SSI & SSDI programs	206	192	91.5	155	80.2	73.5
SSDI only, or in both SSI & SSDI programs	316	305	95.4	239	77.0	73.2
Constructed Disability Status						
Deaf	7	7	100.0	2	30.7	32.1
Mental	235	221	93.5	172	75.6	71.0
Physical	174	171	98.3	142	82.3	80.0
Unknown	11	10	76.6	10	100.0	76.6
Beneficiary's Age (Four Categories)						
18-29 years	57	52	92.5	43	83.6	77.1
30-39 years	85	79	90.0	58	77.8	70.0
40-49 years	148	145	98.0	123	83.6	81.8
50-64 years	137	133	95.8	102	70.2	67.0
Sex						
Male	222	209	92.8	174	79.4	73.6
Female	205	200	97.3	152	77.0	74.5
Hispanicity						
Hispanic	17	15	87.0	14	88.2	76.9
Non-Hispanic/unknown	410	394	95.1	312	78.0	73.9
Race						
White	200	197	98.3	157	76.2	74.9
Black	131	120	87.0	100	85.7	74.7
Unknown	96	92	96.0	69	74.3	71.3
Living Situation						
Living alone	206	192	91.8	152	78.5	72.0
Living with others/unknown	221	217	97.3	174	78.0	75.7
Did the Applicant for Benefits Live in Same Zip Code as Beneficiary?						
No	33	27	79.0	19	73.8	57.1
Yes	216	206	95.1	174	80.9	77.2
No information	178	176	98.3	133	75.6	74.0

TABLE VI.10 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Identity of the Payee with Respect to the Beneficiary						
Beneficiary received beneficiary payments himself or herself	348	334	94.9	266	77.6	73.4
Payee is a family member	62	59	95.0	45	77.1	73.1
Payee is an institution	9	8	89.4	8	100.0	89.4
Other	8	8	100.0	7	96.3	96.7
Type of Claim						
Survivor	16	16	100.0	13	76.1	75.8
Disabled	297	285	94.8	226	78.0	73.6
Unknown	114	108	94.3	87	79.4	74.9
Census Region						
Midwest	31	30	95.0	24	78.9	74.8
Northeast	113	107	94.4	94	79.3	75.1
South	94	94	100.0	72	82.3	81.3
West	189	178	91.2	136	73.9	67.6
Census Division						
East North Central	25	24	93.7	20	82.4	77.1
East South Central	7	7	100.0	7	100.0	100.0
Middle Atlantic	112	106	94.3	94	80.4	76.1
Mountain	7	7	100.0	5	70.8	69.9
New England	1	1	100.0	0	0.0	0.0
Pacific	182	171	90.7	131	74.1	67.5
South Atlantic	38	38	100.0	29	81.5	80.3
West North Central	6	6	100.0	4	66.9	66.7
West South Central	49	49	100.0	36	79.6	78.9
Metropolitan						
Metropolitan areas of 1 million population or more	333	315	92.2	262	80.7	74.7
Metropolitan areas of 250,000 to 999,999 population	54	54	100.0	38	76.3	75.3
Metropolitan areas of less than 250,000 population	18	18	100.0	13	79.4	78.3
Nonmetropolitan areas adjacent to large metropolitan areas	5	5	100.0	4	79.6	80.0
Nonmetropolitan areas adjacent to medium or small metropolitan areas	12	12	100.0	6	50.3	50.0
Nonmetropolitan areas not adjacent to metropolitan areas	5	5	100.0	3	61.0	60.0

Source: NBS, round 3.

TABLE VI.11

WEIGHTED LOCATION AND RESPONSE RATES FOR TICKET PARTICIPANT CROSS-SECTIONAL SAMPLE, PHASE 3 STATES, 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	444	431	97.1	375	86.9	84.4
SSI Only, SSDI Only, or Both SSI and SSDI						
SSI only	149	143	96.0	124	86.7	83.3
SSDI only	183	177	96.7	155	87.5	84.7
Both SSI and SSDI	112	111	99.1	96	86.5	85.7
SSI or SSDI						
SSI only, or in both SSI & SSDI programs	261	254	97.3	220	86.6	84.3
SSDI only, or in both SSI & SSDI programs	295	288	97.6	251	87.1	85.0
Constructed Disability Status						
Deaf	16	16	100.0	13	81.3	81.4
Mental	250	242	96.8	212	87.5	84.8
Physical	173	168	97.1	146	86.9	84.4
Unknown	5	5	100.0	4	79.5	79.7
Beneficiary's Age (Four Categories)						
18-29 years	122	119	97.5	104	87.3	85.2
30-39 years	88	84	95.4	74	88.0	84.1
40-49 years	136	132	97.0	113	85.5	83.1
50-64 years	98	96	97.9	84	87.4	85.7
Sex						
Male	241	234	97.1	206	87.9	85.5
Female	203	197	97.0	169	85.8	83.2
Hispanicity						
Hispanic	30	30	100.0	29	96.6	96.7
Non-Hispanic/unknown	414	401	96.8	346	86.2	83.5
Race						
White	268	263	98.1	236	89.8	88.1
Black	77	73	94.8	64	87.6	83.1
Other/unknown	99	95	95.9	75	78.9	75.8
Living Situation						
Living alone	255	248	97.2	216	87.1	84.7
Living with others/unknown	189	183	96.8	159	86.7	84.1
Did the Applicant for Benefits Live in Same Zip Code as Beneficiary?						
No	34	31	91.1	31	100.0	91.1
Yes	249	246	98.8	215	87.4	86.4
No information	161	154	95.7	129	83.6	80.1
Identity of the Payee with Respect to the Beneficiary						
Beneficiary received beneficiary payments himself or herself	281	272	96.8	238	87.4	84.7
Payee is a family member	133	129	97.0	113	87.6	85.0
Payee is an institution	25	25	100.0	19	75.6	75.7
Other	5	5	100.0	5	100.0	100.0

TABLE VI.11 *(continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Type of Claim						
Survivor	31	29	93.7	24	82.6	77.5
Disabled	280	275	98.2	243	88.3	86.7
Unknown	133	127	95.5	108	85.0	81.3
Census Region						
Midwest	72	69	95.8	64	92.9	89.1
Northeast	28	28	100.0	26	92.7	92.8
South	127	124	97.6	113	91.1	89.0
West	217	210	96.8	172	81.8	79.3
Census Division						
East North Central	48	47	97.9	45	95.8	93.9
East South Central	43	41	95.4	37	90.3	86.1
Middle Atlantic	28	28	100.0	26	92.7	92.8
Mountain	27	26	96.3	19	72.8	70.4
New England	0	0	N/A	0	0	N/A
Pacific	190	184	96.8	153	83.1	80.5
South Atlantic	49	48	98.0	44	91.6	89.7
West North Central	24	22	91.7	19	86.4	79.2
West South Central	35	35	100.0	32	91.6	91.6
Metropolitan						
Metropolitan areas of 1 million population or more	166	159	95.8	132	82.9	79.5
Metropolitan areas of 250,000 to 999,999 population	183	177	96.7	154	87.0	84.2
Metropolitan areas of less than 250,000 population	39	39	100.0	37	95.0	94.9
Nonmetropolitan areas adjacent to large metropolitan areas	9	9	100.0	8	88.4	88.2
Nonmetropolitan areas adjacent to medium or small metropolitan areas	44	44	100.0	42	95.5	95.5
Nonmetropolitan areas not adjacent to metropolitan areas	3	3	100.0	2	65.6	66.7

Source: NBS, round 3.

d. Weighting Classes and Propensity Models for Weight Adjustments

The weight adjustments used in the Ticket Participant Cross-Sectional Sample were based on weighting classes and predicted propensities from a logistic regression model. For the location weighting classes, each class is defined by the levels of the variables associated with locating the sample member. The response probability for a given class c is estimated as follows:

$$\varphi(c) = \frac{sum\ of\ weights\ for\ respondents\ in\ class}{sum\ of\ weights\ for\ all\ sample\ members\ in\ class}$$

For a nonresponding unit within class c, the location adjustment is simply the inverse of this quantity. We calculated the location propensity for traditional Phase 2 participants and all of the cooperation propensities using logistic models. The adjustments for these were simply the inverse of the predicted propensities. The adjusted weight for each sample case is the product of the initial sampling weight and the adjustment factor.

The weighting classes were developed by reviewing Tables VI.6 – VI.11 to determine the variables most closely associated with locating a sample member. The models were developed using the main effects described previously, in addition to two main effects not used in the beneficiary models (the number of times the beneficiary moved in the past five years and the number of changes in the beneficiary‘s phone number in the past five years), plus selected interactions. Interactions to be considered for inclusion in model development were identified using CHAID, as described in the model-fitting section for the Representative Beneficiary Sample.

The primary factors used to calculate the location adjustments are given below, with the potential levels used in the models. Details about how these levels were collapsed for each model are given in Appendix J.

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

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 less than 250,000 residents, (4) beneficiary lived in nonmetropolitan area adjacent to a metropolitan area of 1 million or more, (5) beneficiary lived in nonmetropolitan area adjacent to a metropolitan area of less than 1 million, and (6) beneficiary lived in nonmetropolitan area not adjacent to any metropolitan area.

3. ***GENDER (SEX).*** Possible levels: (1) male, and (2) female.

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

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

6. ***SSI_SSDI.*** Beneficiary status; possible levels: (1) SSI only, (2) SSDI only, or (3) both SSI and SSDI.

7. ***TOC.*** Type of claim; possible levels: (1) survivor claim, (2) disability claim, and (3) type of claim unknown.

8. ***RACE.*** Possible levels: (1) white, (2) black, (3) Asian or Pacific Islander, and (4) not white, black, or Asian/Pacific Islander, or unknown.

A list of the main effects using the variable names listed above for the weight adjustment procedure associated with each payment type-phase combination is provided in Tables VI.12 and VI.13. (An expanded form of Table VI.12, with the specific levels of the main effects for each weighting class or model shown in Tables VI.12 and VI.13, along with parameter estimates and their standard errors, is provided in Appendix J.)

TABLE VI.12

VARIABLES USED IN THE LOCATION ADJUSTMENTS: PHASE 2 TICKET PARTICIPANT CROSS-SECTIONAL SAMPLE

Variables Defining the Location Weighting Class Adjustment for Participants Using the Milestone-Outcome Payment System
Main Effects
SEX
AGECAT
SSI_SSDI
Variables Defining the Location Weighting Class Adjustment for Participants Using the Outcome-Only Payment System
Main Effects
SEX
AGECAT
Variables in the Location Model for Participants Using the Traditional Payment System
Main Effects
LIVING
SSI_SSDI
METRO
TOC
PDZIPSAME

TABLE VI.13

VARIABLES USED IN THE LOCATION ADJUSTMENTS: PHASE 3 TICKET PARTICIPANT CROSS-SECTIONAL SAMPLE

Variables Defining the Location Weighting Class Adjustment for Participants Using the Milestone-Outcome Payment System
Main Effects
PDZIPSAME
SEX
Variables Defining the Location Weighting Class Adjustment for Participants Using the Outcome-Only Payment System
Main Effects
SSI_SSDI
RACE
Variables Defining the Location Weighting Class Adjustment for Participants Using the Traditional Payment System
Main Effects
TOC
RACE

The primary factors in the cooperation models are given below. Only the base variables with all possible levels are given. Some of the base variables below were also listed in the discussion of location adjustments. For these base variables, details about the levels are not given below, since they were provided earlier. (The description of how these levels were collapsed for each model is given in Appendix J.)

1. ***MOVE.*** The number of address changes in the past five years; possible levels: (1) no moves, (2) one move, (3) two or more moves, and (4) information older than five years, or no information.
2. ***DIG.*** Disability diagnostic classification; possible levels: (1) mental disability, (2) physical disability (excluding deaf cases), (3) deaf, and (4) unknown.
3. ***REPREPAYEE.*** The identity of the payee with respect to the beneficiary; possible levels: (1) the beneficiary received payments himself or herself, (2) a family member received benefits on behalf of the beneficiary, and (3) an institution received payments on behalf of the beneficiary, or identity of payee not known.
4. ***PDZIPSAME.*** Whether the beneficiary and the applicant for benefits lived in the same zip code.

5. ***METRO.*** Urbanicity of beneficiary's place of residence.

6. ***GENDER (SEX).***

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, and (9) New England. REGION is based on U.S. Census regions with four levels, which can 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, and (4) Northeast is Middle Atlantic + New England. [54]

8. ***LIVING.*** Beneficiary's living situation.

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

10. ***AGECAT.*** Beneficiary's age category.

11. ***SSI_SSDI.*** Beneficiary status.

12. ***TOC.*** Type of claim.

13. ***RACE***.

14. ***R2LONG.*** Phase 2 longitudinal sample case; possible levels: (1) Phase 2 longitudinal case, and (2) not a Phase 2 longitudinal case.

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

[54] 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.

TABLE VI.14

VARIABLES IN THE COOPERATION LOGISTIC PROPENSITY MODELS:
PHASE 2 TICKET PARTICIPANT CROSS-SECTIONAL SAMPLE

Variables in the Milestone-Outcome Cooperation Model
Main Effects
REPREPAYEE
MOVE
LIVING
SSI_SSDI
DIVISION
TOC
DIG
RACE
AGECAT
Two-Factor Interactions
DIG*REPREPAYEE
TOC*AGECAT
TOC*RACE
DIG*DIVISION
Variables in the Outcome-Only Cooperation Model
Main Effects
RACE
LIVING
DIVISION
SEX
SSI_SSDI
PHONE
PDZIPSAME
AGECAT
R2LONG
METRO
Two-Factor Interactions
REGION*PHONE
R2LONG*AGECAT
R2LONG*REGION
R2LONG*SEX
REGION*AGECAT
AGECAT*R2LONG
AGECAT*SEX
METRO*AGECAT
REGION*SEX
Three-Factor Interactions
R2LONG*SEX*REGION

TABLE VI.14 *(continued)*

Variables in the Traditional Cooperation Model
Main Effects
SSI_SSDI
DIVISION
LIVING
DIG
TOC
AGECAT
REPREPAYEE
METRO
Two-Factor Interactions
DIG*REPREPAYEE
TOC*AGECAT

TABLE VI.15

VARIABLES IN THE COOPERATION LOGISTIC PROPENSITY MODELS: PHASE 3 TICKET PARTICIPANT CROSS-SECTIONAL SAMPLE

Variables in the Milestone-Outcome Cooperation Model
Main Effects
REGION
RACE
SEX
DIG
REPREPAYEE
Two-Factor Interactions
REGION*SEX
Variables in the Outcome-Only Cooperation Model
Main Effects
AGECAT
SEX
REGION
METRO
Two-Factor Interactions
REGION* SEX
Variables in the Traditional Cooperation Model
Main Effects
RACE
REGION
REPREPAYEE
LIVING
Two-Factor Interactions
REPREPAYEE * LIVING

The process for determining which variables would be included in the weighting class adjustments was relatively straightforward. We reviewed Tables VI.6-VI.11 to determine two or three of the variables most closely associated with locating a sample member for each phase-payment type combination, and those variables were used to define the weighting classes. The model-fitting process, however, was considerably more involved. After identifying a smaller pool of main effects and interactions for potential inclusion in the final model, using backward and forward stepwise regressions, a set of models was statistically evaluated to determine the final model. Because the SAS logistic procedure does not incorporate the sampling design, the final selection of the covariates was accomplished using the logistic regression procedure in SUDAAN.

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

The main effects used to calculate the location adjustments (for both weighting classes and models) are summarized in Tables VI.12 and VI.13, and the main effects and interactions in the models for cooperation among located sample members are summarized in Tables VI.14 and VI.15. The R-squared values for the 7 logistic models are given in Table VI.16. The unadjusted R-squared value for the Phase 2 traditional location model was 0.013 (0.054 when rescaled to have a maximum of 1). The unadjusted R-squared value for the nonresponse models ranged from a low of 0.045 (0.084 when rescaled as above) up to 0.127 (0.193 when rescaled). These values are similar to those observed for other response propensity modeling efforts using logistic

regression with design-based sampling weights. The percentages of concordant and discordant pairs, and the p-values for the Hosmer-Lemeshow goodness-of-fit test, are given in Table VI.17.

TABLE VI.16

UNADJUSTED AND ADJUSTED R-SQUARED VALUES FOR LOGISTIC PROPENSITY MODELS IN TICKET PARTICIPANT CROSS-SECTIONAL SAMPLES

Model			Unadjusted R-Squared Value	Adjusted R-Squared Value
Phase	Payment Type	Location or Cooperation		
2	Milestone-Outcome	Location	N/A	N/A
2	Milestone-Outcome	Cooperation	0.076	0.123
2	Outcome-Only	Location	N/A	N/A
2	Outcome-Only	Cooperation	0.127	0.193
2	Traditional	Location	0.013	0.054
2	Traditional	Cooperation	0.052	0.101
3	Milestone-Outcome	Location	N/A	N/A
3	Milestone-Outcome	Cooperation	0.062	0.118
3	Outcome-Only	Location	N/A	N/A
3	Outcome-Only	Cooperation	0.102	0.157
3	Traditional	Location	N/A	N/A
3	Traditional	Cooperation	0.045	0.084

TABLE VI.17

PERCENTAGES OF CONCORDANT AND DISCORDANT PAIRS AND HOSMER-LEMESHOW P-VALUES FOR LOGISTIC PROPENSITY MODELS IN TICKET PARTICIPANT CROSS-SECTIONAL SAMPLES

Model			Percentage Concordant	Percentage Discordant	Hosmer-Lemeshow p-Value
Phase	Payment Type	Location or Cooperation			
2	Milestone-Outcome	Location	N/A	N/A	N/A
2	Milestone-Outcome	Cooperation	66.5	32.2	0.751
2	Outcome-Only	Location	N/A	N/A	N/A
2	Outcome-Only	Cooperation	63.1	36.1	0.179
2	Traditional	Location	63.2	29.2	0.740
2	Traditional	Cooperation	66.8	31.8	0.464
3	Milestone-Outcome	Location	N/A	N/A	N/A
3	Milestone-Outcome	Cooperation	68.5	27.3	0.969
3	Outcome-Only	Location	N/A	N/A	N/A
3	Outcome-Only	Cooperation	65.3	29.2	0.319
3	Traditional	Location	N/A	N/A	N/A
3	Traditional	Cooperation	63.1	28.9	0.389

Although the minimum difference between the percentages of concordant pairs and discordant pairs is 27 percent (Phase 2 outcome-only cooperation model), for the remainder of

the models the difference between these percentages is at least 34 percent. The minimum p-value associated with the Hosmer-Lemeshow goodness-of-fit test is 0.179, indicating no evidence of lack of fit for any of the models.

4. Trimming

After adjustments were applied to the sampling weights, the distribution of weights was reviewed to determine if trimming of the sampling weights was necessary. Because of the wide variation in the magnitude of the weights, due to the use of the composite weights in the outcome-only payment type in Phase 2 and Phase 3 states, and in the milestone-outcome payment type in Phase 2 states, trimming was sometimes necessary to increase the precision of survey estimates. However, we minimized the extent of trimming to reduce the potential for bias in the survey estimates. The design effects due to unequal weighting associated with each of the six phase-payment type combinations before and after trimming, before post-stratification, are presented in Table VI.18. Design effects were calculated separately within trimming strata, which in turn were defined within phase-payment type combinations. In general, the trimming strata were defined according to whether the observation was in the clustered or unclustered sample, and whether the sample was part of the longitudinal or supplemental sample. For unclustered cases, the trimming strata were further subdivided according to whether the sample case was in a PSU or not, and whether the frame used to select the sample value was the round 2 or round 3 frames. The strata within which trimming was employed are given in Table VI.18. If no trimming was employed for a phase-payment type combination, the maximum design effect across all trimming strata is presented. In that instance, the stratum associated with that maximum design effect is not presented since, in most cases, when no trimming is required, the design effects do not differ significantly across trimming strata.

TABLE VI.18

DESIGN EFFECTS DUE TO UNEQUAL WEIGHTS BEFORE AND AFTER TRIMMING, WITHIN TRIMMING STRATA, FOR PHASES AND PAYMENT TYPES IN THE ROUND 3 TICKET PARTICIPANT CROSS-SECTIONAL SAMPLES

Phase and Payment Type	Trimming Stratum in which Trimming Occurred	Design Effect due to Unequal Weights	
Phase 2 Participants		Before Trimming	After Trimming
Milestone-Outcome	Clustered R2 frame supplemental	2.97	2.34
Milestone-Outcome	Clustered R2 frame longitudinal	3.26	2.94
Milestone-Outcome	Clustered R3 frame supplemental	3.47	2.81
Outcome-Only	Clustered R2 frame longitudinal	4.94	2.68
Outcome-Only	Clustered R3 frame supplemental	3.79	3.05
Traditional	No Trimming (three trimming strata)	1.05 (maximum)	1.05(maximum)
Phase 3 Participants		Before Trimming	After Trimming
Milestone-Outcome	No Trimming (one trimming stratum)	1.22	1.22
Outcome-Only	Clustered R3 frame longitudinal	2.73	2.03
Traditional	No Trimming (one trimming stratum)	1.01	1.01

Design Effect due to Unequal Weights = $n\Sigma w^2/(\Sigma w)^2$

5. Post-Stratification

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

D. TICKET PARTICIPANT LONGITUDINAL SAMPLE

For Phase 1 participants who responded in rounds 1, 2, and 3, or only in rounds 1 and 2, the inferential population is defined by the population of Ticket participants at the time of round 1 sampling. Of the 21,477 Phase 1 participants at round 1, 18,100 were participants who used the traditional payment system (84 percent), 2,809 used the milestone-outcome payment system (13

percent) and 568 used the outcome-only payment system (3 percent). In this population, only the participants who used the outcome-only payment system required a supplemental unclustered sample. For Phase 2 participants who responded in rounds 2 and 3, the inferential population is defined by the population of Ticket participants at the time of round 2 sampling. Of the 21,196 Phase 2 participants at round 2, 17,081 were participants who used the traditional payment system (81 percent), 3,208 used the milestone-outcome payment system (15 percent) and 907 used the outcome-only payment system (4 percent). As with the round 3 cross-sectional weights, Phase 2 participants using either the milestone-outcome or outcome-only payment systems required a supplemental unclustered sample. For both the Phase 1 and Phase 2 longitudinal samples, the clustered and unclustered samples were combined using composite weights, the creation of which was described in Section C.

The final sample sizes for the participants longitudinal sample are given in Table VI.19. As stated earlier, the longitudinal samples are determined by the sample selected in round 1 for Phase 1 cases, and by the sample selected in round 2 for Phase 2 cases.

1. Initial Weights

The initial weights were computed based on the probability of selection within the PSU of the released participant sample of Phase 1 cases in round 1 and Phase 2 cases in round 2, and the probability of selection for the PSU. For the unclustered sample for each phase, we computed the initial weights based on the selection probability within the two sampling strata (in one of the PSUs, or not in any PSU).

TABLE VI.19

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

Sampling Strata (Payment System)	Survey Population	Released Sample[55]
Total Phase 1	21,477	1,466
1. Traditional Payment Type	18,100	441
2. Milestone-Outcome Payment Type	2,809	455
3. Outcome-Only Payment Type		570
Clustered Sample	568	123
Unclustered Sample	568	447
In PSUs	123	123
Not in PSUs	445	324
Total Phase 2	21,196	1,350
1. Traditional Payment Type	17,081	437
2. Milestone-Outcome Payment Type		436
Clustered Sample	3,208	216
Unclustered Sample	3,208	220
In PSUs	273	19
Not in PSUs	2,935	201
3. Outcome-Only Payment Type		477
Clustered Sample	907	86
Unclustered Sample	907	391
In PSUs	86	44
Not in PSUs	821	347

Source: Sample allocation and counts computed by MPR.

2. Dual Frame Estimation

To obtain estimates for the paired Ticket Participant Cross-Sectional Samples, it was necessary to combine the clustered and unclustered samples using a paired sample design. As noted earlier, if a potential respondent in the unclustered sample could not be located by the central office, he or she was considered ineligible for field follow-up, and no further attempts were made on that case. However, if a potential respondent was in the clustered sample and could not be located by the central office, the case was sent to the field for additional locating

[55] The final (released) sample size includes participants for whom the number obtained from the original sample design was insufficient for analysis. For Phase 1 participants using the outcome-only payment type, and for Phase 2 participants using the milestone-outcome or outcome-only payment types, a paired sample design was employed, whereby the participants who were in the PSUs potentially could be selected for both samples.

efforts (field follow-up). The paired sample design is the methodology used to combine the samples while accounting for these different rules of field follow-up. This requires the creation of composite weights that can be applied to the combined samples. This same logic is applied to the paired Ticket Participant Longitudinal Samples. The difference, however, is that a sample member may be ineligible for field follow-up in one round, but he or she might be a completed respondent at the central office in the next round. Hence, to create the longitudinal composite weights, different rules were required for defining who was ineligible for field follow-up across rounds. This, of course, depended upon the longitudinal analysis in question. In general, the following rule was applied: if the sample member was ineligible for field follow-up in any of the rounds associated with the longitudinal analysis in question, then the sample member was ineligible for field follow-up in all of the rounds in question. For example, for a longitudinal analysis involving Phase 1 cases where sample members were respondents[56] in rounds 1 or 2, suppose a sample member had a completed interview in round 1, but was ineligible for field follow-up in round 2. For this longitudinal analysis, the sample member would be considered ineligible for field follow-up across the two rounds.

The conceptual framework for composite weights is identical to that given in Section C.2.

3. Nonresponse Adjustment

For the Ticket Participant Longitudinal Sample, we calculated three separate nonresponse adjustments to create three sets of longitudinal weights corresponding to the longitudinal analyses envisioned. In particular, we created two sets of longitudinal weights using the initial Phase 1 longitudinal weights: one for Phase 1 participants who responded in rounds 1 and 2, and

[56] "Respondents" include sample members who had completed interviews, or were ineligible after sample selection (i.e., when surveyed).

the other for Phase 1 participants who responded in all three rounds. In addition, a third set of longitudinal weights was created using the initial Phase 2 longitudinal weights for Phase 2 participants who responded in rounds 2 and 3. As with the Representative Beneficiary and Ticket Participant Cross-Sectional Samples, the nonresponse adjustments were accomplished in two stages for each set of weights: one stage for the sample members who could not be located, and another stage for those who, once located, refused to respond. All adjustments were calculated using logistic regression propensity models. We applied separate adjustments for each payment type, applying the nonresponse adjustments to the composite weights for the unclustered and clustered samples, where appropriate. This resulted in a total of 18 logistic regression models. These models were fitted in the same way as the adjustment models for the Representative Beneficiary Sample, described in Section B.2, and the adjustment models for the Ticket Participant Cross-Sectional Sample, described in Section C.3. The specific covariates for each of the 18 logistic models are described in subsequent sections.

a. Coding of Survey Dispositions

For cross-sectional estimates, identifying completed cases and calculating response rates was straightforward once sample members were categorized into one of the following four groups:

1. Eligible respondents.
2. Ineligible respondents (sample members who were ineligible after sample selection, including deceased, sample members in the military or incarcerated, sample members living outside of the United States, and other ineligibles).
3. Located nonrespondents (including active or passive refusals, and language barrier situations).
4. Unlocated sample members (sample members who could not be located using either central office tracing procedures or in-field searches).

However, for longitudinal estimates, rules were necessary to assign sample members to one of the four categories for all rounds covered by the longitudinal weight, based upon the categories into which they were classified for each round. For example, for longitudinal analyses involving Phase 1 cases for all three rounds, a complete case was defined as someone who responded in all three rounds. In the list of rules for classifying respondents into the four categories for longitudinal analyses, a ―round-specific" respondent or nonrespondent indicates that the sample member did or did not respond in a particular round. The definition of ―across rounds" depends upon the longitudinal analysis in question. In particular, it means either (1) ―across rounds 1 and 2" for longitudinal analyses involving Phase 1 cases across rounds 1 and 2; or (2) ―across rounds 1, 2, and 3" for longitudinal analyses involving Phase 1 cases across rounds 1, 2, and 3; or (3) ―across rounds 2 and 3" for longitudinal analyses involving Phase 2 cases across rounds 2 and 3. The list of rules is given below:

1. ***Longitudinal Eligible Respondents.*** Eligible respondents across rounds.
2. ***Longitudinal Ineligible Respondents.*** Sample members who were either ineligible or eligible respondents across rounds, and were ineligible respondents in at least one round (they were not considered longitudinal eligible respondents).
3. ***Longitudinal Unlocated Sample Members.*** Sample members who were either ineligible respondents, eligible respondents, or were not located across rounds, and were not located in at least one round (they were neither longitudinal eligible nor longitudinal ineligible respondents).
4. ***Longitudinal Located Nonrespondents.*** Sample members who were located and did not respond in at least one round, and in all other rounds were either eligible respondents, ineligible respondents, or unlocated sample members.

b. Response Rates

We would expect the response rates for the Ticket Participant Longitudinal Sample to be lower than those obtained by the cross-sectional sample, since response is defined across rounds, and sample attrition would result in fewer responses in each round. When response is defined across all three rounds for Phase 1 cases, the response rate is 62.8 percent. This rate is a product

of the weighted location rate (92.1 percent) and the weighted completion rate among located sample members (68.2 percent). The same Phase 1 initial longitudinal weights (adjusted for the released sample) are used to calculate the response rate for longitudinal analyses across the first two rounds (70.5 percent), which is higher because response is required only across two rounds, instead of three. This is a product of the weighted location rate (93.6 percent) and the weighted completion rate among located sample members (75.3 percent). Finally, the response rate for Phase 2 cases for analyses across rounds 2 and 3 is 71.7 percent, the product of the weighted location rate (91.4 percent) and the weighted completion rate among located cases (78.4 percent).

c. Factors Related to Location and Response

Tables VI.20 – VI.28 provide information for selected factors associated with locating a sample member within each payment type associated with each set of longitudinal weights. The tables include unweighted counts of all sample members, counts of located sample members across rounds, and counts of the sample members for whom a completed interview was obtained, or who were determined to be ineligible when surveyed across each round. The tables also include the weighted location rate, the weighted completion rate among located sample members, and the weighted overall completion rate for these factors, which helped inform the decision about the final set of variables used in the nonresponse adjustment models.

d. Propensity Models for Weight Adjustments

The weight adjustments used in the Ticket Participant Longitudinal Sample were based on predicted propensities from a logistic regression model. For the location and cooperation weight adjustments, we used logistic models to estimate the propensity for a sample member to be located and cooperate across rounds. The inverse of the propensity score was used as the adjustment factor. The adjusted weight for each sample case is the product of the initial sampling weight and the adjustment factor.

TABLE VI.20

WEIGHTED LOCATION AND RESPONSE RATES FOR TICKET PARTICIPANT LONGITUDINAL SAMPLE, PHASE 1 STATES, ROUNDS 1 AND 2 RESPONDENTS, MILESTONE-OUTCOME PAYMENT SYSTEM, BY SELECTED CHARACTERISTICS

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
All	455	398	90.3	300	76.8	69.6
SSI Only, SSDI Only, or Both SSI and SSDI						
SSI only	166	144	88.4	109	78.2	69.6
SSDI only	197	178	92.7	128	72.5	67.4
Both SSI and SSDI	92	76	88.3	63	84.2	74.3
SSI or SSDI						
SSI only, or in both SSI & SSDI programs	258	220	88.3	172	80.6	71.5
SSDI only, or in both SSI & SSDI programs	289	254	91.3	191	76.1	69.6
Constructed Disability Status						
Deaf	6	5	84.6	4	76.3	67.8
Mental	257	224	90.5	179	80.6	72.9
Physical	179	159	91.1	110	71.9	65.9
Unknown	13	10	76.7	7	66.1	52.3
Beneficiary's Age (Four Categories)						
18-29 years	79	61	80.2	49	79.0	63.6
30-39 years	114	104	93.4	81	76.9	71.9
40-49 years	122	112	93.8	79	73.1	69.0
50-64 years	140	121	89.6	91	79.1	71.1
Sex						
Male	331	290	89.8	214	73.5	66.0
Female	124	108	91.7	86	84.6	78.2
Hispanicity						
Hispanic	19	12	64.9	11	89.4	59.0
Non-Hispanic/unknown	436	386	91.1	289	76.5	69.9
Race						
White	131	117	93.1	88	80.2	75.1
Black	115	103	91.4	85	85.7	78.5
Other/unknown	209	178	87.5	127	69.0	60.2
Living Situation						
Living alone	264	228	89.5	179	80.6	72.4
Living with others/unknown	191	170	91.4	121	72.2	66.2
Did the Applicant for Benefits Live in Same Zip Code as Beneficiary?						
No	43	32	79.6	23	76.3	61.3
Yes	205	179	90.3	145	83.1	75.4
No information	207	187	92.6	132	70.9	65.7
Identity of the Payee with Respect to the Beneficiary						
Beneficiary received beneficiary payments himself or herself	351	307	90.1	231	76.4	69.1
Payee is a family member	80	70	91.5	55	80.3	73.7
Payee is an institution	18	15	86.9	8	58.4	48.7
Other	6	6	100.0	6	100.0	100.0

TABLE VI.20 (*continued)*

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
Changes in Telephone Number						
No changes in last 5 years	314	277	91.3	210	77.6	70.9
One or more changes in last 5 years	12	9	76.3	9	100.0	76.3
No information/other	129	112	88.7	81	72.7	65.0
Number of Moves in Last 5 Years						
No moves last 5 years	130	109	86.5	87	79.1	68.8
One or more moves in last 5 years	11	6	59.5	5	77.9	45.6
No information on number of moves	314	283	92.7	208	76.0	70.6
Type of Claim						
Survivor	36	31	90.3	24	80.2	72.3
Disabled	262	232	91.7	174	75.8	69.7
Unknown	157	135	87.5	102	78.0	68.7
Census Region						
Midwest	75	67	92.0	55	82.1	75.2
Northeast	124	106	87.9	76	74.3	65.5
South	148	133	93.3	99	76.8	72.0
West	108	92	86.0	70	75.3	64.9
Census Division						
East North Central	73	65	91.0	54	86.2	78.5
East South Central	0	0	N/A	0	0	N/A
Middle Atlantic	77	66	89.1	49	77.7	69.6
Mountain	91	78	86.8	59	74.5	65.0
New England	47	40	85.8	27	67.9	58.0
Pacific	17	14	82.4	11	79.4	64.7
South Atlantic	147	132	92.9	98	75.6	70.7
West North Central	2	2	100.0	1	50.0	50.0
West South Central	1	1	100.0	1	100.0	100.0
Metropolitan						
Metropolitan areas of 1 million population or more	404	349	87.0	263	75.2	65.5
Metropolitan areas of 250,000 to 999,999 population	36	34	96.3	25	81.1	78.2
Metropolitan areas of less than 250,000 population	11	11	100.0	8	73.3	73.1
Nonmetropolitan areas adjacent to large metropolitan areas	0	0	N/A	0	0	N/A
Nonmetropolitan areas adjacent to medium or small metropolitan areas	4	4	100.0	4	100.0	100.0
Nonmetropolitan areas not adjacent to metropolitan areas	0	0	N/A	0	0	N/A

Source: NBS, round 3.

TABLE VI.21

WEIGHTED LOCATION AND RESPONSE RATES FOR TICKET PARTICIPANT LONGITUDINAL SAMPLE, PHASE 1 STATES, ROUNDS 1 AND 2 RESPONDENTS, OUTCOME-ONLY PAYMENT SYSTEM, BY SELECTED CHARACTERISTICS

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
All	489	457	90.9	326	72.0	64.7
SSI Only, SSDI Only, or Both SSI and SSDI						
SSI only	94	84	82.4	57	70.8	56.8
SSDI only	319	301	92.2	219	74.6	68.0
Both SSI and SSDI	76	72	95.2	50	62.7	60.2
SSI or SSDI						
SSI only, or in both SSI & SSDI programs	170	156	88.4	107	67.0	58.4
SSDI only, or in both SSI & SSDI programs	395	373	92.8	269	72.3	66.5
Constructed Disability Status						
Deaf	4	3	77.3	1	29.0	22.7
Mental	258	241	90.4	164	68.7	61.3
Physical	214	201	91.9	155	77.6	70.9
Unknown	13	12	90.7	6	64.6	51.4
Beneficiary's Age (Four Categories)						
18-29 years	35	35	100.0	25	70.5	70.3
30-39 years	106	93	86.1	69	73.0	61.5
40-49 years	167	155	91.0	107	67.4	61.2
50-64 years	181	174	92.2	125	76.5	69.5
Sex						
Male	297	276	91.8	193	70.5	64.4
Female	192	181	89.3	133	74.5	65.2
Hispanicity						
Hispanic	13	12	90.6	5	42.3	39.9
Non-Hispanic/unknown	476	445	90.9	321	72.6	65.3
Race						
White	264	254	93.0	193	76.6	70.3
Black	65	56	87.2	39	66.5	58.6
Other/unknown	160	147	88.4	94	65.4	57.1
Living Situation						
Living alone	183	171	90.3	122	71.2	63.5
Living with others/unknown	306	286	91.2	204	72.5	65.5
Did the Applicant for Benefits Live in Same Zip Code as Beneficiary?						
No	39	36	93.9	21	57.2	54.8
Yes	191	181	91.3	140	74.0	66.7
No information	259	240	90.2	165	72.5	64.6

TABLE VI.21 (*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	409	387	91.0	281	71.7	64.7
Payee is a family member	58	50	86.5	28	59.4	47.3
Payee is an institution	11	11	100.0	8	87.6	88.7
Other	11	9	89.5	9	100.0	89.5
Changes in Telephone Number						
No changes in last 5 years	416	395	92.2	281	70.4	64.5
One or more changes in last 5 years	11	11	100.0	11	100.0	100.0
No information/other	62	51	81.6	34	77.9	62.0
Number of Moves in Last 5 Years						
No moves last 5 years	118	108	89.7	80	67.2	59.8
One or more moves in last 5 years	8	8	100.0	7	85.6	85.4
No information on number of moves	363	341	91.1	239	73.2	65.9
Type of Claim						
Survivor	19	17	89.6	10	62.8	56.3
Disabled	384	361	92.6	260	71.7	66.2
Unknown	86	79	83.2	56	74.9	59.9
Census Region						
Midwest	98	89	90.4	66	74.6	67.0
Northeast	144	137	91.9	98	73.4	68.0
South	125	123	98.5	75	56.3	56.3
West	122	108	82.9	87	83.9	68.5
Census Division						
East North Central	85	76	88.4	58	77.8	68.2
East South Central	0	0	N/A	0	0	N/A
Middle Atlantic	53	50	88.5	30	66.9	59.2
Mountain	57	54	93.7	41	83.9	76.7
New England	91	87	93.8	68	77.0	72.9
Pacific	65	54	75.0	46	84.0	62.5
South Atlantic	121	119	98.4	73	56.5	56.4
West North Central	13	13	100.0	8	60.7	61.5
West South Central	4	4	100.0	2	49.2	50.0
Metropolitan						
Metropolitan areas of 1 million population or more	300	273	86.3	195	73.3	62.2
Metropolitan areas of 250,000 to 999,999 population	72	70	97.5	51	68.8	67.5
Metropolitan areas of less than 250,000 population	42	41	97.6	26	65.0	63.6
Nonmetropolitan areas adjacent to large metropolitan areas	4	4	100.0	4	100.0	100.0
Nonmetropolitan areas adjacent to medium or small metropolitan areas	33	32	97.0	23	72.0	69.7
Nonmetropolitan areas not adjacent to metropolitan areas	38	37	97.4	27	72.8	71.1

Source: NBS, round 3.

TABLE VI.22

WEIGHTED LOCATION AND RESPONSE RATES FOR TICKET PARTICIPANT LONGITUDINAL SAMPLE, PHASE 1 STATES, ROUNDS 1 AND 2 RESPONDENTS, 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	441	414	94.2	310	75.2	70.8
SSI Only, SSDI Only, or Both SSI and SSDI						
SSI only	118	107	90.9	77	71.7	65.3
SSDI only	214	204	95.7	149	73.8	70.6
Both SSI and SSDI	109	103	94.6	84	81.4	77.0
SSI or SSDI						
SSI only, or in both SSI & SSDI programs	227	210	92.7	161	76.4	70.9
SSDI only, or in both SSI & SSDI programs	323	307	95.4	233	76.5	72.8
Constructed Disability Status						
Deaf	21	20	95.8	13	65.1	62.4
Mental	221	209	94.6	148	71.5	67.4
Physical	196	183	94.0	148	80.7	76.0
Unknown	3	2	59.0	1	62.4	38.9
Beneficiary's Age (Four Categories)						
18-29 years	77	73	95.2	51	69.4	65.5
30-39 years	93	84	91.4	65	77.3	71.0
40-49 years	140	129	92.3	97	77.1	71.1
50-64 years	131	128	97.5	97	75.3	73.4
Sex						
Male	242	225	93.4	175	77.7	72.8
Female	199	189	95.2	135	72.0	68.2
Hispanicity						
Hispanic	24	22	91.8	20	92.2	84.0
Non-Hispanic/unknown	417	392	94.3	290	74.2	70.0
Race						
White	226	213	94.8	162	76.2	72.3
Black	115	105	91.4	79	75.5	68.9
Other/unknown	100	96	95.9	69	72.5	69.3
Living Situation						
Living alone	222	207	93.5	160	76.8	71.9
Living with others/unknown	219	207	94.9	150	73.5	69.5
Did the Applicant for Benefits Live in Same Zip Code as Beneficiary?						
No	33	29	89.7	22	77.1	68.7
Yes	224	211	94.1	167	79.5	74.9
No information	184	174	94.9	121	69.6	66.1

TABLE VI.22 (*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	307	289	94.4	213	74.2	70.1
Payee is a family member	99	92	93.6	66	71.7	66.6
Payee is an institution	26	25	95.2	23	91.1	86.6
Other	9	8	89.4	8	100.0	89.4
Changes in Telephone Number						
No changes in last 5 years	316	299	94.8	228	76.8	72.7
One or more changes in last 5 years	6	6	100.0	5	83.9	84.1
No information/other	119	109	92.3	77	70.6	65.1
Number of Moves in Last 5 Years						
No moves last 5 years	137	128	93.5	102	80.3	75.2
One or more moves in last 5 years	6	6	100.0	6	100.0	100.0
No information on number of moves	298	280	94.3	202	72.3	68.1
Type of Claim						
Survivor	43	42	97.6	31	75.2	72.6
Disabled	286	272	95.5	206	76.2	72.6
Unknown	112	100	89.6	73	72.6	65.4
Census Region						
Midwest	166	158	95.2	126	79.9	75.9
Northeast	127	118	93.4	80	68.4	64.1
South	119	113	94.8	83	73.8	70.0
West	29	25	86.2	21	85.3	73.2
Census Division						
East North Central	160	152	95.0	121	79.8	75.6
East South Central	0	0	N/A	0	0	N/A
Middle Atlantic	116	108	93.5	72	67.5	63.3
Mountain	23	20	87.2	16	81.4	69.9
New England	11	10	91.7	8	81.7	75.1
Pacific	6	5	83.3	5	100.0	83.3
South Atlantic	117	111	94.7	81	73.3	69.5
West North Central	6	6	100.0	5	83.0	83.0
West South Central	2	2	100.0	2	100.0	100.0
Metropolitan						
Metropolitan areas of 1 million population or more	206	190	92.8	133	70.3	64.9
Metropolitan areas of 250,000 to 999,999 population	137	130	94.8	104	80.3	76.1
Metropolitan areas of less than 250,000 population	64	62	97.0	46	74.5	72.2
Nonmetropolitan areas adjacent to large metropolitan areas	0	0	N/A	0	0	N/A
Nonmetropolitan areas adjacent to medium or small metropolitan areas	19	17	89.4	13	75.3	68.6
Nonmetropolitan areas not adjacent to metropolitan areas	15	15	100.0	14	93.2	93.3

Source: NBS, round 3.

TABLE VI.23

WEIGHTED LOCATION AND RESPONSE RATES FOR TICKET PARTICIPANT LONGITUDINAL SAMPLE, PHASE 1 STATES, ROUNDS 1, 2, AND 3 RESPONDENTS, MILESTONE-OUTCOME PAYMENT SYSTEM, BY SELECTED CHARACTERISTICS

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
All	455	390	89.0	264	70.9	63.4
SSI Only, SSDI Only, or Both SSI and SSDI						
SSI only	166	139	85.7	95	72.0	62.0
SSDI only	197	176	92.0	111	66.0	61.1
Both SSI and SSDI	92	75	87.4	58	80.0	70.6
SSI or SSDI						
SSI only, or in both SSI & SSDI programs	258	214	86.4	153	75.2	65.5
SSDI only, or in both SSI & SSDI programs	289	251	90.5	169	70.4	64.1
Constructed Disability Status						
Deaf	6	5	84.6	4	80.7	67.8
Mental	257	219	89.0	155	73.9	66.0
Physical	179	157	90.2	98	66.2	60.3
Unknown	13	9	69.3	7	74.2	52.3
Beneficiary's Age (Four Categories)						
18-29 years	79	61	80.2	47	75.7	60.5
30-39 years	114	103	92.8	66	67.3	62.6
40-49 years	122	109	91.9	69	67.4	63.0
50-64 years	140	117	87.4	82	75.1	66.1
Sex						
Male	331	284	88.3	191	68.2	60.5
Female	124	106	90.5	73	77.6	70.6
Hispanicity						
Hispanic	19	11	59.0	10	93.7	55.1
Non-Hispanic/unknown	436	379	89.9	254	70.4	63.7
Race						
White	131	116	92.6	76	73.6	68.8
Black	115	102	90.7	76	79.1	71.9
Other/unknown	209	172	85.0	112	64.3	54.5
Living Situation						
Living alone	264	222	87.6	155	73.6	65.0
Living with others/unknown	191	168	90.7	109	67.5	61.6
Did the Applicant for Benefits Live in Same Zip Code as Beneficiary?						
No	43	29	73.8	18	68.3	51.9
Yes	205	176	89.1	128	77.3	69.3
No information	207	185	91.9	118	65.3	60.2

TABLE VI.23 (*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	351	303	89.1	203	70.0	63.0
Payee is a family member	80	68	89.7	49	75.4	67.7
Payee is an institution	18	14	82.5	8	62.1	48.7
Other	6	5	83.7	4	81.8	67.4
Changes in Telephone Number						
No changes in last 5 years	314	271	89.9	183	71.5	64.8
One or more changes in last 5 years	12	9	76.3	9	100.0	76.3
No information/other	129	110	87.3	72	66.6	58.2
Number of Moves in Last 5 Years						
No moves last 5 years	130	106	84.6	76	72.5	62.2
One or more moves in last 5 years	11	6	59.5	4	67.2	38.6
No information on number of moves	314	278	91.5	184	70.4	64.7
Type of Claim						
Survivor	36	31	90.3	22	75.5	68.2
Disabled	262	229	90.9	152	69.4	63.7
Unknown	157	130	84.6	90	72.7	61.7
Census Region						
Midwest	75	66	91.1	50	77.2	70.2
Northeast	124	104	86.4	64	65.7	57.1
South	148	130	92.0	88	72.0	67.0
West	108	90	84.3	62	69.0	58.0
Census Division						
East North Central	73	64	89.9	49	80.8	72.9
East South Central	0	0	N/A	0	0	N/A
Middle Atlantic	77	65	87.8	40	67.4	59.9
Mountain	91	76	84.8	52	68.5	57.9
New England	47	39	83.7	24	62.6	51.9
Pacific	17	14	82.4	10	71.8	58.8
South Atlantic	147	129	91.6	87	70.6	65.4
West North Central	2	2	100.0	1	49.8	50.0
West South Central	1	1	100.0	1	100.0	100.0
Metropolitan						
Metropolitan areas of 1 million population or more	404	341	85.1	229	67.5	57.5
Metropolitan areas of 250,000 to 999,999 population	36	34	96.3	23	77.6	74.8
Metropolitan areas of less than 250,000 population	11	11	100.0	8	73.2	73.1
Nonmetropolitan areas adjacent to large metropolitan areas	0	0	N/A	0	0	N/A
Nonmetropolitan areas adjacent to medium or small metropolitan areas	4	4	100.0	4	100.0	100.0
Nonmetropolitan areas not adjacent to metropolitan areas	0	0	N/A	0	0	N/A

Source: NBS, round 3.

TABLE VI.24

WEIGHTED LOCATION AND RESPONSE RATES FOR TICKET PARTICIPANT LONGITUDINAL SAMPLE, PHASE 1 STATES, ROUNDS 1, 2, AND 3 RESPONDENTS, OUTCOME-ONLY PAYMENT SYSTEM, BY SELECTED CHARACTERISTICS

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
All	453	419	88.1	275	67.1	58.3
SSI Only, SSDI Only, or Both SSI and SSDI						
SSI only	82	72	82.5	47	71.2	57.5
SSDI only	299	283	90.7	187	66.2	59.6
Both SSI and SSDI	72	64	84.0	41	65.8	54.0
SSI or SSDI						
SSI only, or in both SSI & SSDI programs	154	136	83.2	88	68.8	56.0
SSDI only, or in both SSI & SSDI programs	371	347	89.5	228	66.2	58.5
Constructed Disability Status						
Deaf	4	3	77.2	1	28.8	22.8
Mental	234	217	86.2	132	66.5	56.1
Physical	204	188	90.3	136	69.6	62.8
Unknown	11	11	100.0	6	52.8	47.2
Beneficiary's Age (Four Categories)						
18-29 years	33	33	100.0	22	66.9	65.0
30-39 years	95	84	81.1	54	63.3	49.7
40-49 years	154	140	87.5	88	66.6	58.1
50-64 years	171	162	91.1	111	69.7	62.6
Sex						
Male	276	251	89.0	165	68.5	60.2
Female	177	168	86.6	110	64.5	55.1
Hispanicity						
Hispanic	13	12	90.5	5	43.9	39.9
Non-Hispanic/unknown	440	407	88.1	270	67.6	58.7
Race						
White	247	233	89.4	162	71.3	62.9
Black	57	50	81.1	33	73.9	58.5
Other/unknown	149	136	88.9	80	57.1	50.4
Living Situation						
Living alone	168	152	85.8	101	71.6	60.1
Living with others/unknown	285	267	89.6	174	64.2	57.2
Did the Applicant for Benefits Live in Same Zip Code as Beneficiary?						
No	36	32	87.3	18	55.8	48.9
Yes	175	160	86.8	111	69.9	59.8
No information	242	227	89.2	146	66.4	58.4

TABLE VI.24 (*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	383	356	86.9	237	64.2	55.7
Payee is a family member	52	47	89.8	26	59.1	49.3
Payee is an institution	9	9	100.0	5	86.3	83.0
Other	9	7	94.3	7	100.0	94.3
Changes in Telephone Number						
No changes in last 5 years	393	369	88.6	239	64.8	56.7
One or more changes in last 5 years	9	9	100.0	9	100.0	100.0
No information/other	51	41	84.3	27	74.5	62.4
Number of Moves in Last 5 Years						
No moves last 5 years	108	97	82.6	62	60.8	49.4
One or more moves in last 5 years	7	7	100.0	5	67.8	65.9
No information on number of moves	338	315	89.5	208	68.9	60.8
Type of Claim						
Survivor	18	16	89.1	9	60.4	53.7
Disabled	360	336	89.2	220	65.8	58.0
Unknown	75	67	83.2	46	74.1	60.4
Census Region						
Midwest	89	79	87.2	53	69.4	59.3
Northeast	135	127	90.4	79	68.8	61.2
South	113	108	92.1	63	52.1	49.4
West	116	105	82.8	80	77.1	63.0
Census Division						
East North Central	77	67	84.5	46	71.8	59.5
East South Central	0	0	N/A	0	0	N/A
Middle Atlantic	49	44	88.7	21	68.6	59.0
Mountain	55	52	93.8	37	78.7	73.1
New England	86	83	91.6	58	68.9	62.7
Pacific	61	53	75.0	43	75.8	55.9
South Atlantic	109	104	91.8	61	52.2	49.4
West North Central	12	12	100.0	7	59.0	58.3
West South Central	4	4	100.0	2	48.5	50.0
Metropolitan						
Metropolitan areas of 1 million population or more	280	252	83.9	170	67.1	55.5
Metropolitan areas of 250,000 to 999,999 population	66	62	89.5	40	72.2	63.8
Metropolitan areas of less than 250,000 population	36	36	100.0	20	57.3	57.4
Nonmetropolitan areas adjacent to large Metropolitan areas	4	4	100.0	3	75.2	75.0
Nonmetropolitan areas adjacent to medium or small metropolitan areas	31	30	96.8	19	63.4	61.3
Nonmetropolitan areas not adjacent to Metropolitan areas	36	35	97.2	23	65.6	63.9

Source: NBS, round 3.

TABLE VI.25

WEIGHTED LOCATION AND RESPONSE RATES FOR TICKET PARTICIPANT LONGITUDINAL SAMPLE, PHASE 1 STATES, ROUNDS 1, 2, AND 3 RESPONDENTS, 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	441	406	92.7	274	67.8	62.8
SSI Only, SSDI Only, or Both SSI and SSDI						
SSI only	118	105	89.3	71	67.2	60.1
SSDI only	214	199	94.0	128	65.6	61.4
Both SSI and SSDI	109	102	94.1	75	72.5	68.6
SSI or SSDI						
SSI only, or in both SSI & SSDI programs	227	207	91.6	146	69.8	64.2
SSDI only, or in both SSI & SSDI programs	323	301	94.0	203	67.9	63.9
Constructed Disability Status						
Deaf	21	20	95.8	10	49.6	47.5
Mental	221	205	93.1	133	64.9	60.6
Physical	196	179	92.5	130	73.3	67.5
Unknown	3	2	59.0	1	59.5	38.9
Beneficiary's Age (Four Categories)						
18-29 years	77	71	93.2	48	66.8	62.1
30-39 years	93	84	91.4	57	67.4	62.0
40-49 years	140	126	90.3	84	68.6	62.0
50-64 years	131	125	96.0	85	67.6	64.8
Sex						
Male	242	221	92.0	152	68.9	63.4
Female	199	185	93.6	122	66.3	62.1
Hispanicity						
Hispanic	24	21	87.4	18	85.2	74.2
Non-Hispanic/unknown	417	385	93.0	256	66.8	62.2
Race						
White	226	211	94.3	142	67.9	64.0
Black	115	102	89.1	72	70.4	63.0
Other/unknown	100	93	93.2	60	64.3	59.9
Living Situation						
Living alone	222	204	92.3	145	70.2	65.1
Living with others/unknown	219	202	93.1	129	65.2	60.5
Did the Applicant for Benefits Live in Same Zip Code as Beneficiary?						
No	33	29	89.7	20	70.0	63.0
Yes	224	207	92.7	148	71.5	66.6
No information	184	170	93.3	106	62.7	58.2

TABLE VI.25 (*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	307	283	93.0	185	66.0	61.4
Payee is a family member	99	91	92.6	61	66.2	61.4
Payee is an institution	26	24	91.8	20	81.7	74.8
Other	9	8	89.4	8	100.0	89.4
Changes in Telephone Number						
No changes in last 5 years	316	293	93.4	200	68.6	64.1
One or more changes in last 5 years	6	5	83.6	3	60.7	51.4
No information/other	119	108	91.5	71	66.0	60.2
Number of Moves in Last 5 Years						
No moves last 5 years	137	126	92.3	93	73.8	68.5
One or more moves in last 5 years	6	5	83.7	4	77.9	67.5
No information on number of moves	298	275	93.1	177	64.8	60.2
Type of Claim						
Survivor	43	42	97.6	29	70.3	68.3
Disabled	286	266	93.9	178	67.3	63.2
Unknown	112	98	87.8	67	67.9	59.9
Census Region						
Midwest	166	157	94.6	113	71.8	68.0
Northeast	127	117	92.8	72	62.4	58.0
South	119	108	91.4	71	66.4	60.6
West	29	24	83.1	18	74.6	62.6
Census Division						
East North Central	160	151	94.3	109	72.0	68.1
East South Central	0	0	N/A	0	0	N/A
Middle Atlantic	116	108	93.5	65	61.4	57.5
Mountain	23	19	83.0	14	73.2	61.2
New England	11	9	83.4	7	78.2	65.1
Pacific	6	5	83.3	4	79.6	66.7
South Atlantic	117	106	91.2	69	65.7	59.9
West North Central	6	6	100.0	4	66.8	65.9
West South Central	2	2	100.0	2	100.0	100.0
Metropolitan						
Metropolitan areas of 1 million population or more	206	185	90.7	120	65.1	59.0
Metropolitan areas of 250,000 to 999,999 population	137	127	93.3	89	70.3	65.7
Metropolitan areas of less than 250,000 population	64	62	97.0	38	61.6	59.6
Nonmetropolitan areas adjacent to large metropolitan areas	0	0	N/A	0	0	N/A
Nonmetropolitan areas adjacent to medium or small metropolitan areas	19	17	89.4	13	76.5	68.6
Nonmetropolitan areas not adjacent to metropolitan areas	15	15	100.0	14	93.2	93.3

Source: NBS, round 3.

TABLE VI.26

WEIGHTED LOCATION AND RESPONSE RATES FOR TICKET PARTICIPANT LONGITUDINAL SAMPLE, PHASE 2 STATES, ROUNDS 2 AND 3 RESPONDENTS, MILESTONE-OUTCOME PAYMENT SYSTEM, BY SELECTED CHARACTERISTICS

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
All	389	363	92.0	275	76.8	69.8
SSI Only, SSDI Only, or Both SSI and SSDI						
SSI only	147	132	85.9	99	67.3	56.4
SSDI only	153	146	96.1	114	84.8	80.6
Both SSI and SSDI	89	85	94.4	62	76.3	71.2
SSI or SSDI						
SSI only, or in both SSI & SSDI programs	236	217	88.9	161	70.6	61.6
SSDI only, or in both SSI & SSDI programs	242	231	95.5	176	82.1	77.6
Constructed Disability Status						
Deaf	16	12	91.8	4	29.1	25.0
Mental	227	206	87.9	154	77.0	66.0
Physical	138	137	98.4	111	83.5	82.3
Unknown	8	8	100.0	6	49.0	51.0
Beneficiary's Age (Four Categories)						
18-29 years	117	108	85.4	80	74.7	61.7
30-39 years	86	78	92.8	54	73.3	67.8
40-49 years	102	95	95.4	74	76.5	72.4
50-64 years	84	82	99.2	67	84.4	83.2
Sex						
Male	204	191	91.9	144	76.4	70.0
Female	185	172	92.1	131	77.3	69.7
Hispanicity						
Hispanic	4	3	83.8	3	100.0	83.8
Non-Hispanic/unknown	385	360	92.0	272	76.7	69.7
Race						
White	211	199	91.8	149	72.9	66.2
Black	141	130	92.3	97	79.6	72.2
Other/unknown	37	34	91.6	29	90.6	82.4
Living Situation						
Living alone	226	206	87.6	156	72.7	62.3
Living with others/unknown	163	157	97.5	119	81.8	79.0
Did the Applicant for Benefits Live in Same Zip Code as Beneficiary?						
No	38	32	85.7	24	75.1	63.4
Yes	202	189	90.2	143	74.2	65.7
No information	149	142	95.5	108	80.4	76.1

TABLE VI.26 (*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	228	214	93.8	176	83.4	77.1
Payee is a family member	122	114	90.2	77	70.4	62.1
Payee is an institution	30	27	89.4	17	63.3	56.3
Other	9	8	86.1	5	74.3	63.5
Changes in Telephone Number						
No changes in last 5 years	263	250	91.7	200	83.2	75.4
One or more changes in last 5 years	12	9	84.8	7	79.0	64.3
No information/other	114	104	93.4	68	60.1	55.9
Number of Moves in Last 5 Years						
No moves last 5 years	123	112	83.6	93	81.6	66.6
One or more moves in last 5 years	11	9	85.5	6	64.4	48.6
No information on number of moves	255	242	96.0	176	75.1	72.0
Type of Claim						
Survivor	51	49	98.5	33	73.5	72.6
Disabled	201	192	95.1	151	84.0	78.9
Unknown	137	122	85.0	91	66.0	54.5
Census Region						
Midwest	131	118	91.2	91	79.7	72.5
Northeast	22	22	100.0	17	79.4	77.9
South	226	213	91.6	159	73.7	66.0
West	10	10	100.0	8	79.9	80.0
Census Division						
East North Central	109	98	92.1	78	81.5	75.3
East South Central	40	40	100.0	30	80.2	78.5
Middle Atlantic	10	10	100.0	7	71.2	70.5
Mountain	9	9	100.0	7	77.7	77.8
New England	12	12	100.0	10	87.9	86.8
Pacific	1	1	100.0	1	100.0	100.0
South Atlantic	24	24	100.0	20	82.1	82.3
West North Central	22	20	86.9	13	69.5	58.6
West South Central	162	149	85.6	109	68.5	56.1
Metropolitan						
Metropolitan areas of 1 million population or more	250	231	89.6	181	76.4	67.5
Metropolitan areas of 250,000 to 999,999 population	37	36	98.0	25	69.8	66.7
Metropolitan areas of less than 250,000 population	40	39	98.0	28	81.6	80.0
Nonmetropolitan areas adjacent to large metropolitan areas	32	28	86.2	22	81.3	69.5
Nonmetropolitan areas adjacent to medium or small metropolitan areas	14	14	100.0	7	51.6	50.0
Nonmetropolitan areas not adjacent to metropolitan areas	16	15	94.1	12	83.2	76.4

Source: NBS, round 3.

TABLE VI.27

WEIGHTED LOCATION AND RESPONSE RATES FOR TICKET PARTICIPANT LONGITUDINAL SAMPLE, PHASE 2 STATES, ROUNDS 2 AND 3 RESPONDENTS, OUTCOME-ONLY PAYMENT SYSTEM, BY SELECTED CHARACTERISTICS

	Sample	Located Sample		Response Among Located Sample		Overall Respondents
	Count	Count	Location Rate	Count	Response Rate	Response Rate
All	414	396	93.6	265	66.9	62.4
SSI Only, SSDI Only, or Both SSI and SSDI						
SSI only	62	58	94.6	39	75.6	70.1
SSDI only	296	285	96.5	186	63.0	61.0
Both SSI and SSDI	56	53	79.2	40	75.9	59.6
SSI or SSDI						
SSI only, or in both SSI & SSDI programs	118	111	87.4	79	75.7	65.2
SSDI only, or in both SSI & SSDI programs	352	338	93.4	226	65.0	60.8
Constructed Disability Status						
Deaf	11	11	100.0	6	55.9	54.8
Mental	210	201	96.3	138	66.1	63.8
Physical	191	182	90.3	120	68.5	61.3
Unknown	2	2	100.0	1	50.4	50.0
Beneficiary's Age (Four Categories)						
18-29 years	45	43	95.5	27	64.0	60.4
30-39 years	101	96	94.9	58	61.3	57.6
40-49 years	158	151	89.7	101	71.8	63.8
50-64 years	110	106	98.0	79	65.4	65.0
Sex						
Male	218	209	95.5	130	62.6	59.8
Female	196	187	91.7	135	71.1	64.8
Hispanicity						
Hispanic	9	9	100.0	7	74.0	73.7
Non-Hispanic/unknown	405	387	93.5	258	66.7	62.2
Race						
White	278	267	96.3	176	65.6	63.4
Black	83	80	96.2	51	63.7	60.8
Other/unknown	53	49	75.8	38	79.7	59.1
Living Situation						
Living alone	131	124	88.4	87	74.6	65.0
Living with others/unknown	283	272	96.4	178	63.0	60.9
Did the Applicant for Benefits Live in Same Zip Code as Beneficiary?						
No	33	32	98.9	24	70.1	70.4
Yes	150	143	89.2	100	71.4	63.2
No information	231	221	95.6	141	63.9	60.8

TABLE VI.27 (*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	334	320	93.2	215	67.5	62.7
Payee is a family member	56	54	96.3	33	58.2	55.3
Payee is an institution	13	12	96.5	11	95.8	93.2
Other	11	10	89.2	6	57.4	52.7
Changes in Telephone Number						
No changes in last 5 years	341	328	93.4	216	64.3	60.0
One or more changes in last 5 years	5	4	80.0	4	100.0	80.0
No information/other	68	64	95.4	45	76.8	72.5
Number of Moves in Last 5 Years						
No moves last 5 years	89	87	98.4	59	69.3	68.0
One or more moves in last 5 years	5	4	80.0	4	100.0	80.0
No information on number of moves	320	305	92.6	202	65.8	60.7
Type of Claim						
Survivor	32	31	96.7	22	71.5	68.8
Disabled	324	310	92.8	205	64.1	59.6
Unknown	58	55	95.7	38	77.3	72.7
Census Region						
Midwest	77	71	86.6	53	73.3	62.4
Northeast	229	220	95.2	141	63.1	60.4
South	94	92	98.6	62	67.1	66.4
West	14	13	92.9	9	69.8	64.3
Census Division						
East North Central	40	37	95.7	25	52.6	46.2
East South Central	37	37	100.0	24	64.8	64.9
Middle Atlantic	53	48	90.8	30	61.3	55.3
Mountain	11	10	90.9	7	70.4	63.6
New England	176	172	96.5	111	63.6	62.0
Pacific	3	3	100.0	2	67.6	66.7
South Atlantic	31	30	96.9	20	65.7	63.8
West North Central	37	34	80.9	28	90.3	72.3
West South Central	26	25	98.7	18	72.3	71.6
Metropolitan						
Metropolitan areas of 1 million population or more	153	144	95.2	101	70.5	66.6
Metropolitan areas of 250,000 to 999,999 population	159	156	96.8	98	62.3	60.5
Metropolitan areas of less than 250,000 population	31	29	93.5	21	72.6	67.7
Nonmetropolitan areas adjacent to large metropolitan areas	13	11	73.1	8	92.2	66.6
Nonmetropolitan areas adjacent to medium or small metropolitan areas	37	36	97.3	24	66.9	64.9
Nonmetropolitan areas not adjacent to metropolitan areas	21	20	97.2	13	40.3	40.1

Source: NBS, round 3.

TABLE VI.28

WEIGHTED LOCATION AND RESPONSE RATES FOR TICKET PARTICIPANT LONGITUDINAL SAMPLE, PHASE 2 STATES, ROUNDS 2 AND 3 RESPONDENTS, 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	437	397	91.2	316	79.5	72.6
SSI Only, SSDI Only, or both SSI and SSDI						
SSI only	107	101	94.9	74	72.2	68.3
SSDI only	221	198	89.8	160	81.1	72.9
Both SSI and SSDI	109	98	90.6	82	83.4	76.0
SSI or SSDI						
SSI only, or in both SSI & SSDI programs	216	199	92.7	156	77.8	72.2
SSDI only, or in both SSI & SSDI programs	330	296	90.0	242	81.9	73.9
Constructed Disability Status						
Deaf	18	14	81.9	9	63.7	52.2
Mental	241	221	91.7	179	80.4	73.8
Physical	171	155	91.2	125	81.3	74.1
Unknown	7	7	100.0	3	40.6	43.4
Beneficiary's Age (Four Categories)						
18-29 years	123	114	93.4	89	78.0	72.6
30-39 years	96	82	85.8	58	70.0	60.4
40-49 years	123	114	92.7	95	83.8	77.6
50-64 years	95	87	92.1	74	85.2	78.5
Sex						
Male	219	197	90.3	154	78.6	70.8
Female	218	200	92.1	162	80.3	74.3
Hispanicity						
Hispanic	4	3	73.8	1	42.4	27.0
Non-Hispanic/unknown	433	394	91.4	315	79.8	73.0
Race						
White	286	265	93.0	211	79.6	74.1
Black	99	87	88.2	69	78.4	69.1
Other/unknown	52	45	87.0	36	80.7	70.4
Living Situation						
Living alone	218	201	92.7	160	79.2	73.6
Living with others/unknown	219	196	89.7	156	79.8	71.5
Did the Applicant for Benefits Live in Same Zip Code as Beneficiary?						
No	41	37	90.8	27	73.4	66.7
Yes	222	200	90.6	162	80.7	73.0
No information	174	160	92.0	127	79.3	73.3

TABLE VI.28 (*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	284	255	90.1	203	79.9	72.0
Payee is a family member	114	107	94.4	88	81.8	77.2
Payee is an institution	25	23	92.8	20	86.7	80.7
Other	14	12	86.3	5	39.6	34.6
Changes in Telephone Number						
No changes in last 5 years	316	283	89.9	227	80.2	72.4
One or more changes in last 5 years	8	8	100.0	6	74.3	74.0
No information/other	113	106	94.2	83	77.7	72.9
Number of Moves in Last 5 Years						
No moves last 5 years	124	109	88.8	87	80.1	71.3
One or more moves in last 5 years	7	7	100.0	7	100.0	100.0
No information on number of moves	306	281	92.0	222	78.7	72.4
Type of Claim						
Survivor	48	44	91.7	36	83.2	76.0
Disabled	288	258	90.0	209	80.9	73.1
Unknown	101	95	94.6	71	73.6	69.4
Census Region						
Midwest	217	200	92.3	154	76.8	71.2
Northeast	24	22	91.0	18	83.4	76.1
South	161	143	89.5	117	81.5	73.0
West	35	32	91.4	27	85.0	77.1
Census Division						
East North Central	202	185	91.6	141	75.9	69.9
East South Central	42	33	81.2	25	74.3	59.7
Middle Atlantic	9	9	100.0	6	66.4	66.7
Mountain	35	32	91.4	27	85.0	77.1
New England	15	13	87.1	12	91.6	80.3
Pacific	0	0	N/A	0	0	N/A
South Atlantic	72	67	93.2	57	84.8	79.1
West North Central	15	15	100.0	13	86.3	85.9
West South Central	47	43	91.0	35	82.3	74.8
Metropolitan						
Metropolitan areas of 1 million population or more	194	180	92.7	134	74.2	69.1
Metropolitan areas of 250,000 to 999,999 population	55	49	89.2	38	77.3	69.0
Metropolitan areas of less than 250,000 population	70	58	84.5	51	87.9	74.7
Nonmetropolitan areas adjacent to large metropolitan areas	44	42	96.1	36	85.2	81.6
Nonmetropolitan areas adjacent to medium or small metropolitan areas	12	11	91.9	9	78.4	71.9
Nonmetropolitan areas not adjacent to metropolitan areas	62	57	91.9	48	84.3	77.4

Source: NBS, round 3.

The models were developed with the same main effects used in the development of the nonresponse models for the beneficiary weights and Ticket participant cross-sectional weights, plus selected interactions. Interactions to be considered for inclusion in model development were identified using CHAID, as described previously.

The primary factors used in both the location and cooperation models for the longitudinal weights are given below, along with potential levels used in the models. (Details about how these levels were collapsed are given in Appendix K.)

1. ***MOVE.*** The number of address changes in the past five years; possible levels: (1) no moves, (2) one move, (3) two or more moves, and (4) information older than five years, or no information.
2. ***DIG.*** Disability diagnostic classification; possible levels: (1) mental disability, (2) physical disability (excluding deaf cases), (3) deaf, and (4) unknown.
3. ***REPREPAYEE.*** The identity of the payee with respect to the beneficiary; possible levels: (1) the beneficiary received payments himself or herself, (2) a family member received benefits on behalf of the beneficiary, and (3) an institution received payments on behalf of the beneficiary, or identity of payee not known
4. ***PDZIPSAME.*** Whether the beneficiary and the applicant for benefits lived in the same zip code; possible levels: (1) beneficiary and applicant lived in the same zip code, (2) beneficiary and applicant lived in different zip codes, and (3) information unknown.
5. ***METRO.*** Urbanicity of beneficiary‘s place of residence; possible levels: (1) beneficiary lived in metropolitan area of 1 million or more residents, (2) beneficiary lived in metropolitan area of 250,000 to 999,999, (3) beneficiary lived in metropolitan area of less than 250,000 residents, (4) beneficiary lived in nonmetropolitan area adjacent to a metropolitan area of 1 million or more, (5) beneficiary lived in nonmetropolitan area adjacent to a metropolitan area of less than 1 million, and (6) beneficiary lived in nonmetropolitan area not adjacent to any metropolitan area.
6. ***GENDER (SEX).*** Possible levels: (1) male, and (2) female.
7. ***REGION 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, and (9) New England. REGION is based on U.S. Census regions, with four levels, which can be collapsed from the nine levels of DIVISION: (1) West is Pacific and Mountain, (2) Midwest is East North Central and West North Central, (3) South is

East South Central, West South Central, and South Atlantic, and (4) Northeast is Middle Atlantic and New England.[57]

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

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

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

11. ***SSI_SSDI.*** Beneficiary status; possible levels: (1) SSI only, (2) SSDI only, or (3) both SSI and SSDI.

12. ***TOC.*** Type of claim; possible levels: (1) survivor claim, (2) disability claim, and (3) type of claim unknown.

13. ***RACE.*** Race; possible levels: (1) white, (2) black, (3) Asian or Pacific Islander, and (4) not white, black, or Asian/Pacific Islander, or unknown.

Various interactions among these variables were also included in the model for locating the sample member. A list of the main effects using variable names listed above, as well as interactions, is provided in Tables VI.29 – VI.34. An expanded form of Tables VI.29 – VI.34, with the specific levels of the main effects for each model and the interactions shown in Tables VI.29 – VI.34, along with parameter estimates and their standard errors, is provided in Appendix K. In Appendix K, the variables are followed by suffixes representing the collapsing of the base variable's levels unique to each model. This follows the procedure used in the cross-sectional models. These suffixes are not shown in Tables VI.29 – VI.34.

[57] Many of the 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.

TABLE VI.29

VARIABLES USED IN THE LOCATION LOGISTIC PROPENSITY MODELS: TICKET PARTICIPANT LONGITUDINAL SAMPLE, PHASE 1 STATES, ROUNDS 1 AND 2 RESPONDENTS

Variables in the Location Model for Participants Using Milestone-Outcome Payment System
Main Effects
METRO
AGECAT
PDZIPSAME
TOC
REGION
Two-Factor Interactions
REGION*TOC
Variables in the Location Model for Participants Using Outcome-Only Payment System
Main Effects
AGECAT
GENDER (SEX)
SSI_SSDI
DIVISION
REPREPAYEE
LIVING
METRO
Two-Factor Interactions
DIVISION*AGECAT
Variables in the Location Model for Participants Using Traditional Payment System
Main Effects
GENDER (SEX)
RACE
TOC
DIG
REGION
SSI_SSDI
AGECAT
REPREPAYEE
Two-Factor Interactions
SEX * TOC
SSI_SSDI* REPREPAYEE
SSI_SSDI* DIG
RACE* TOC

TABLE VI.30

VARIABLES USED IN THE LOCATION LOGISTIC PROPENSITY MODELS: TICKET PARTICIPANT LONGITUDINAL SAMPLE, PHASE 1 STATES, ROUNDS 1, 2, AND 3 RESPONDENTS

Variables in the Location Model for Participants Using Milestone-Outcome Payment System
Main Effects
MOVE
PDZIPSAME
METRO
RACE
AGECAT
Variables in the Location Model for Participants Using Outcome-Only Payment System
Main Effects
MOVE
LIVING
RACE
AGECAT
DIVISION
SSI_SSDI
METRO
Two-Factor Interactions
MOVE*SSI_SSDI
DIVISION*RACE
SSI_SSDI*RACE
AGECAT*DIVISION
Variables in the Location Model for Participants Using Traditional Payment System
Main Effects
GENDER (SEX)
SSI_SSDI
REGION
RACE
AGECAT
Two-Factor Interactions
RACE*SEX

TABLE VI.31

VARIABLES USED IN THE LOCATION LOGISTIC PROPENSITY MODELS: TICKET PARTICIPANT LONGITUDINAL SAMPLE, PHASE 2 STATES, ROUNDS 2 AND 3 RESPONDENTS

Variables in the Location Model for Participants Using Milestone-Outcome Payment System
Main Effects
AGECAT
TOC
GENDER (SEX)
REGION
DIG
PHONE
MOVE
REPREPAYEE
SSI_SSDI
LIVING
Two-Factor Interactions
MOVE*AGECAT
SEX*AGECAT
MOVE*REGION
DIG*REPREPAYEE
Variables in the Location Model for Participants Using Outcome-Only Payment System
Main Effects
AGECAT
GENDER (SEX)
REGION
MOVE
RACE
Two-Factor Interactions
RACE*REGION
REGION*SEX
Variables in the Location Model for Participants Using Traditional Payment System
Main Effects
DIVISION
RACE
PHONE
METRO
AGECAT
REPREPAYEE
Two-Factor Interactions
REPREPAYEE*PHONE
AGECAT*PHONE
METRO*AGECAT

TABLE VI.32

VARIABLES USED IN THE COOPERATION LOGISTIC PROPENSITY MODELS: TICKET PARTICIPANT LONGITUDINAL SAMPLE, PHASE 1 STATES, ROUNDS 1 AND 2 RESPONDENTS

Variables in the Cooperation Model for Participants Using Milestone-Outcome Payment System
Main Effects
PHONE
RACE
PDZIPSAME
DIG
Two-Factor Interactions
PDZIPSAME* DIG
Variables in the Cooperation Model Participants Using Outcome-Only Payment System
Main Effects
GENDER (SEX)
MOVE
REGION
DIG
REPREPAYEE
LIVING
SSI_SSDI
RACE
PDZIPSAME
Two-Factor Interactions
REGION *RACE
SEX*REGION
MOVE*RACE
DIG*REGION
Variables in the Cooperation Model for Participants Using Traditional Payment Systems
Main Effects
PDZIPSAME
METRO
REPREPAYEE
DIG
REGION
SSI_SSDI
Two-Factor Interactions
SSI_SSDI*DIG

TABLE VI.33

VARIABLES USED IN THE COOPERATION LOGISTIC PROPENSITY MODELS: TICKET PARTICIPANT LONGITUDINAL SAMPLE, PHASE 1 STATES, ROUNDS 1, 2, AND 3 RESPONDENTS

Variables in the Cooperation Model for Participants Using Milestone-Outcome Payment System
Main Effects
MOVE
PDZIPSAME
SSI_SSDI
METRO
RACE
AGECAT
DIG
Two-Factor Interactions
SSI_SSDI*DIG
AGECAT*PDZIPSAME
Variables in the Cooperation Model for Participants Using Outcome-Only Payment System
Main Effects
MOVE
RACE
AGECAT
PDZIPSAME
SSI_SSDI
REGION
DIG
LIVING
Two-Factor Interactions
MOVE*REGION
DIG*AGECAT
Variables in the Cooperation Model for Participants Using Traditional Payment System
Main Effects
DIG
PDZIPSAME
AGECAT
METRO
REGION
REPREPAYEE
RACE
Two-Factor Interactions
REGION* PDZIPSAME
DIG*AGECAT
REGION*RACE

TABLE VI.34

VARIABLES USED IN THE COOPERATION LOGISTIC PROPENSITY MODELS: TICKET PARTICIPANT LONGITUDINAL SAMPLE, PHASE 2 STATES, ROUNDS 2 AND 3 RESPONDENTS

Variables in the Cooperation Model for Participants Using Milestone-Outcome Payment System
Main Effects
AGECAT
REPREPAYEE
DIG
SSI_SSDI
PHONE
RACE
MOVE
TOC
Two-Factor Interactions
AGECAT*RACE
AGECAT*PHONE
REPREPAYEE*RACE
RACE*TOC
SSI_SSDI*REPREPAYEE
SSI_SSDI*MOVE
REPREPAYEE*TOC
Variables in the Cooperation Model for Participants Using Outcome-Only Payment System
Main Effects
GENDER (SEX)
METRO
DIVISION
DIG
PDZIPSAME
Two-Factor Interactions
DIG* PDZIPSAME
Variables in the Cooperation Model for Participants Using Traditional Payment Systems
Main Effects
GENDER (SEX)
PDZIPSAME
REPREPAYEE
DIG
REGION
SSI_SSDI
AGECAT
METRO
Two-Factor Interactions
SEX*PDZIPSAME
SSI_SSDI*REPREPAYEE
SSI_SSDI*DIG
SSI_SSDI*REGION
REGION*AGECAT

The process for identifying the final model was exactly the same as for the Representative Beneficiary Sample (described in Section B.2) and the Ticket Participant Cross-Sectional Sample (described in Section C.3). The R-squared values for the 18 logistic models are given in Table VI.35. Overall, the unadjusted R-squared values ranged from a low of 0.039 (0.097 when rescaled to have a maximum of 1) to a high of 0.211 (0.449 when rescaled to have a maximum of 1). However, the models tend to be better with the Phase 2 than the Phase 1 longitudinal sample. Among Phase 1 cases, the models are better when looking only at rounds 1 and 2 respondents rather than rounds 1, 2, and 3 respondents. These values are similar to those observed for other response propensity efforts using logistic regression and design-based sampling weights.

TABLE VI.35

UNADJUSTED AND ADJUSTED R-SQUARED VALUES FOR LOGISTIC PROPENSITY MODELS IN THE TICKET PARTICIPANT LONGITUDINAL SAMPLE

Model				
Respondent in Round	Payment Type	Location or Cooperation	Unadjusted R-Squared Value	Adjusted R-Squared Value
1 and 2	Milestone-Outcome	Location	0.067	0.142
1 and 2	Milestone-Outcome	Cooperation	0.077	0.116
1 and 2	Outcome-Only	Location	0.141	0.309
1 and 2	Outcome-Only	Cooperation	0.192	0.277
1 and 2	Traditional	Location	0.082	0.227
1 and 2	Traditional	Cooperation	0.078	0.116
1, 2, and 3	Milestone-Outcome	Location	0.078	0.156
1, 2, and 3	Milestone-Outcome	Cooperation	0.100	0.143
1, 2, and 3	Outcome-Only	Location	0.141	0.271
1, 2, and 3	Outcome-Only	Cooperation	0.138	0.192
1, 2, and 3	Traditional	Location	0.039	0.097
1, 2, and 3	Traditional	Cooperation	0.091	0.127
2 and 3	Milestone-Outcome	Location	0.192	0.449
2 and 3	Milestone-Outcome	Cooperation	0.211	0.319
2 and 3	Outcome-Only	Location	0.146	0.386
2 and 3	Outcome-Only	Cooperation	0.089	0.124
2 and 3	Traditional	Location	0.063	0.140
2 and 3	Traditional	Cooperation	0.155	0.243

The percentages of concordant and discordant pairs, and the p-values for the Hosmer-Lemeshow goodness-of-fit test, are given in Table VI.36. Although the minimum difference between the percentages of concordant pairs and discordant pairs is 20.3 percent (Phase 2 cooperation model for participants using the outcome-only payment system), for the remainder of models the difference between these percentages is at least 26.5 percent, and all but 3 have differences greater than 30 percent. The minimum p-value associated with the Hosmer-Lemeshow goodness-of-fit test is 0.209, indicating no evidence of lack of fit for any of the models.

TABLE VI.36

PERCENTAGES OF CONCORDANT AND DISCORDANT PAIRS AND HOSMER-LEMESHOW P-VALUES FOR LOGISTIC PROPENSITY MODELS IN THE TICKET PARTICIPANT LONGITUDINAL SAMPLE

Model					
Respondent in Round	Payment Type	Location or Cooperation	Percent Concordant	Percent Discordant	Hosmer-Lemeshow p-value
1 and 2	Milestone-Outcome	Location	58.6	26.6	0.948
1 and 2	Milestone-Outcome	Cooperation	60.9	33.5	0.983
1 and 2	Outcome-Only	Location	78.7	18.7	0.631
1 and 2	Outcome-Only	Cooperation	74.1	25.0	0.425
1 and 2	Traditional	Location	77.2	20.5	0.938
1 and 2	Traditional	Cooperation	65.8	29.1	0.570
1, 2, and 3	Milestone-Outcome	Location	60.9	29.7	0.850
1, 2, and 3	Milestone-Outcome	Cooperation	61.9	35.4	0.575
1, 2, and 3	Outcome-Only	Location	68.4	28.7	0.629
1, 2, and 3	Outcome-Only	Cooperation	64.8	33.5	0.247
1, 2, and 3	Traditional	Location	65.1	28.3	0.714
1, 2, and 3	Traditional	Cooperation	68.2	29.9	0.217
2 and 3	Milestone-Outcome	Location	74.9	22.6	0.209
2 and 3	Milestone-Outcome	Cooperation	70.1	28.7	0.268
2 and 3	Outcome-Only	Location	63.0	31.4	0.314
2 and 3	Outcome-Only	Cooperation	55.7	35.4	0.496
2 and 3	Traditional	Location	71.8	22.8	0.945
2 and 3	Traditional	Cooperation	76.9	22.4	0.839

4. Trimming

After adjustments were applied to the longitudinal weights, the distribution of weights was reviewed to determine if trimming of the longitudinal weights was necessary. Because of the wide variation in the magnitude of the weights due to the use of the composite weights in some situations, trimming was sometimes necessary to increase the precision of survey estimates. However, we minimized the extent of trimming to reduce the potential for bias in the survey estimates. The design effects due to unequal weights associated with each type of longitudinal weight and payment type before and after trimming, before post-stratification, are presented in Table VI.37. Design effects were calculated separately within trimming strata, which in turn were defined within longitudinal weight type and payment type. In general, weights in the clustered sample were larger than those in the unclustered sample. Therefore, the trimming strata were defined according to whether the observation was in the clustered or unclustered sample. For unclustered cases, the trimming strata were further subdivided according to whether the sample case was in a PSU or not. The strata within which trimming was employed are given in Table VI.37. As with the trimming of cross-sectional weights, if no trimming was employed for a longitudinal weight type and payment type, we would have presented the maximum design effect due to unequal weights across all trimming strata. However, it always turned out that only one trimming stratum was used in these instances. In light of this, if no trimming was employed, the design effect due to unequal weighting within that single trimming stratum was presented.

TABLE VI.37

DESIGN EFFECTS DUE TO UNEQUAL WEIGHTS BEFORE AND AFTER TRIMMING, WITHIN TRIMMING STRATA, FOR EACH TYPE OF LONGITUDINAL WEIGHT AND PAYMENT TYPE

Longitudinal Weight Type and Payment Type		Trimming Stratum Where Trimming Occurred	Design Effect due to Unequal Weights	
Respondent in Round	Payment Type		Before Trimming	After Trimming
1 and 2	Milestone-Outcome	No Trimming	1.66	1.66
1 and 2	Outcome-Only	Clustered Sample	3.37	2.67
1 and 2	Traditional	No Trimming	1.06	1.06
1, 2, and 3	Milestone-Outcome	No Trimming	1.54	1.54
1, 2, and 3	Outcome-Only	Clustered Sample	3.80	3.12
1, 2, and 3	Traditional	No Trimming	1.10	1.10
2 and 3	Milestone-Outcome	Clustered Sample	4.16	2.88
2 and 3	Outcome-Only	Clustered Sample	3.69	2.88
2 and 3	Traditional	No Trimming	1.15	1.15

5. Post-Stratification

After the nonresponse adjustment and trimming, the weights were post-stratified to the population age and gender totals for each payment type obtained from the SSA sampling frame as it was defined when the phase was first rolled out. This sampling frame included all SSI or SSDI beneficiaries for each payment type within the Ticket Participant population. For Phase 1 cases, the weight totals were defined by the round 1 sampling frame, which included only data from Phase 1 states. For Phase 2 cases, the weight totals were defined by the round 2 sampling frame for Phase 2 states only. The distributions of weights within each phase and payment type combination were rechecked to determine if more weight trimming was necessary. No additional weight trimming was required.

VII. IMPUTATIONS

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

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

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

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

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

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

preceding or following a sample member with missing data was randomly selected as the donor with equal probability. The hot-deck procedure was therefore 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. Because the extent of missing values was very low, only a few donors were used more than once.

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

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

defining an imputation class, the donor also was missing the value for that variable. This strategy was employed if there were large numbers of donors and recipients missing the covariate in question. In the second instance, for a given recipient, a variable was used 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 would depend upon 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 these cases, imputation classes were collapsed, and the imputation was redone using the collapsed classes. The strategy for collapsing classes required a ranking of the variables used to define the imputation class with regard to each variable‘s relationship to the variable requiring imputation. If a number of covariates were used to aid in the imputation of a given variable, the covariates less closely related to the variable requiring imputation were more likely to have levels collapsed than more important covariates in the imputation. In addition, variables with many levels also were more likely to have levels collapsed. In general, if more than a very small number of imputation classes required collapsing, then one or more variables were dropped from the definition of the imputation class, and the imputation procedure was rerun.

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

For some of the imputed variables in the data set, the number of missing responses does not match the number of imputed responses. Often, these variables correspond to questions that follow a filter question. For example, question I33 asks if the respondent has difficulty climbing

10 steps, and if the response is –yes," the follow-up question (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. This caused the count of the actual number of imputed responses to be greater than the number of missing or invalid responses.

A. NBS IMPUTATIONS OF SPECIFIC VARIABLES

Included below in several tables is information about how imputation was employed in the NBS. The tables include the imputed variable names and a brief description of each. The tables also include the methods of imputation, total number of missing responses, the number of respondents eligible for the question, and the percentage of responses imputed. This information was recorded 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 a previous round). In most cases, the logical assignments were done using imputed values.[58] Therefore, the distinction between –logically assigned" and –constructed from other variables with imputed values" is somewhat opaque. In

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

general, if a logical assignment was done for variables corresponding directly to questionnaire questions, the flag was set to 1. For variables constructed from these variables (constructed variables are prefixed with a –C_"), the flag was set to 6. In this instance, one or more of the component variables in the constructed variable were imputed.

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

1. Section L: Race and Ethnicity

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

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

TABLE VII.1

RACE AND ETHNICITY IMPUTATIONS

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
L1_i	Hispanic/Latino Ethnic Origins	3 imputations from SSA's administrative data, 2 longitudinal imputations, 57 imputations from hot deck	62	6,605	0.94
C_Race_i	Race	15 longitudinal imputations, 93 imputations from SSA's administrative data, 126 imputations from hot deck	234	6,605	3.54

Source: NBS, round 3.

Respondents could choose from five race categories: white, black/African American, Asian, Hawaiian/Pacific Islander, and Native American/American Indian. Respondents were allowed to select more than one of these categories to identify themselves (as prescribed by the Office of Management and Budget). The final race variable on which imputation was applied had six categories, with a separate category for respondents reporting multiple races. Although the SSA administrative data did not have a category for multiple races, respondents with race information in the SSA files were categorized according to four of the five categories above (Hawaiian/Pacific Islanders were included with the respondents reporting Asian). Respondents who did not answer the race question but did have race information in the SSA files were categorized into one of the four categories. This resulted in misclassification of respondents with extant SSA administrative data who did not answer the race question in the survey, but would have identified themselves in the survey as multiple race or Hawaiian/Pacific Islander. However, we assumed that the number of respondents of this kind was small, so that misclassification was not a major problem. As with the Hispanic indicator, for respondents that still had missing data, race was imputed using a hot deck with imputation classes defined by the

zip code of each sample member, and Hispanicity as a sorting variable. If the respondent was a longitudinal case, then the imputed value from round 1 was used.

2. Section B: Disability Status Variables and Work Indicator

Table VII.2 describes five imputed variables that pertain to the sample member's disability status and an indicator of whether the respondent was currently working. These imputed variables include three that collapse and recode primary diagnosis codes from the International Classification of Diseases, Ninth Revision (ICD-9) in three different ways: C_MainConBodyGroup_i, which corresponds to the collapsing done in Table II.2; C_MainConDiagGrp_i; and C_MainConColDiagGrp_i. Additional disability status variables include age when the disability was first diagnosed (C_DisAge_i); and an indicator of childhood or adult onset of the disability (C_AdultChildOnset_i). A fourth variable with collapsed primary diagnosis codes also was imputed, with levels further collapsed from C_MainConDiagGrp_i. This variable (C_MainConImput_i) is not included in Table VII.2 because it was not released to the final file, but it was used in subsequent imputations as a classing variable. As with race and ethnicity, the age when the disability was first diagnosed cannot change from one round to the next. For 14 missing values among longitudinal cases, this age variable was obtained from round 1 data. All missing values for C_AdultChildOnset_i were "logically assigned" using the imputed values from C_DisAge_i, the age-of-onset variable. In addition, Section B contains a question asking whether the respondent was currently working (B24_i). This is a gate question for all of the work status variables in Section C.

TABLE VII.2

DISABILITY STATUS IMPUTATIONS

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
C_MainConDiagGrp_i	Primary diagnosis group	56 hot deck	56	5,906	0.95
C_MainConColDiagGrp_i	Main condition diagnosis group collapsed	56 hot deck	56	5,906	0.95
C_MainConBodyGroup_i	Main condition body group	56 hot deck	56	5,906	0.95
C_Disage_i	Age at onset of disability	167 hot deck; 43 from longitudinal data	210	6,605	3.18
C_Adultchild_onset_i	Adult/child onset of disability	19 logical	19	6,605	0.29
B24_i	Currently working	6 hot deck	6	6,605	0.09

Source: NBS, round 3.

All of the variables in Section B used an indicator of whether the onset of the disability was in childhood or adulthood, as well as age and gender, to define imputation classes. One of the collapsed condition code variables, C_MainConImput_i was also used as a classing variable for disability age and the work indicator. Additional classing variables were used that were specific to the variable being imputed.

3. Section C: Current Jobs Variables

Several questions in the NBS asked respondents about current employment. In Section C, these questions were asked only of respondents who indicated in question B24 that they were currently working. These include salary (C_MainCurJobHrPay_i, C_MainCurJobMnthPay_i, and C_TotCurJobMnthPay_i); usual hours worked at the job or jobs (C8_1_i, C_TotCurWkHrs_i, and C_TotCurHrMnth_i); the number of places the respondent was

employed (C1_i); and job description of the place of main employment (C2_1_1d_i). These variables are identified in Table VII.3.

TABLE VII.3

CURRENT JOBS IMPUTATIONS

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
C1_i	Number of current jobs	2 hot deck	2	1,739	0.12
C2_1_1d_i	Main current job SOC code to one digit	8 hot deck[a]	8	1,739	0.46
C8_1_i	Hours per week usually worked at current main job	26 hot deck[b]; 1 imputed by distributional assumptions	27	1,739	1.55
C_TotCurWkHrs_i	Total weekly hours at all current jobs	26 hot deck[c], 6 constructed from imputed variables	32	1,739	1.84
C_TotCurHrMnth_i	Total hours per month at all current jobs	32 constructed from imputed variables	32	1,739	1.84
C_MainCurJobHrPay_i	Hourly pay at current main job	3 logical, 183 constructed from imputed variables	186	1,739	10.70
C_MainCurJobMnthPay_i	Monthly pay at current main job	10 logical, 23 imputed by distributional assumptions, 160 constructed from imputed variables	193	1,739	11.10
C_TotCurMnthPay_i	Total monthly salary all current jobs	31 logical, 160 hot deck, 11 constructed from imputed variables	202	1,739	11.62

Source: NBS, round 3.

[a]Imputations for current job variables include 2 cases coded as "don't know" or "refused" in B24, which were imputed as currently working in B24_i.

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

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

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

Other variables had values imputed by using the distribution of a variable related to the variable at hand. For example, if the take-home monthly pay of the respondent's current main job was not missing but the gross monthly pay (C_MainCurJobMnthPay_i) of this job was missing, then the relationship between gross monthly and take-home monthly pay among respondents missing neither variable was used to determine the appropriate value for gross monthly pay. In particular, a random draw was selected from the observed distribution of relative taxes, where "relative tax" is defined as the proportion of imputed gross monthly pay for 22 cases with missing data for C_MainCurJobMnthPay. As Table VII.3 indicates, hot-deck imputations were applied to only four of the jobs variables: C1_i, C2_1_1d_i, C8_1_i, and C_TotCurMnthPay_i. For these variables, the collapsed condition code variable and level of education were used as classing variables. Additional classing and sorting variables specific to each variable also were used.

4. Section I: Health Status Variables

A total of 56 health status variables where imputations were applied are in Section I of the National Beneficiary Survey questionnaire. The 56 imputed variables in this section, and the methods of imputation used in each case, are identified in Tables VII.4 and VII.5. These items

cover a range of topics, from the respondent‘s general health to more specific questions on the instrumental activities of daily living (IADLs) and activities of daily living (ADLs), and other health and coping indicators. Also included in this section are a series of questions pertaining to the respondent‘s use of illicit drugs and alcohol.

TABLE VII.4

HEALTH STATUS IMPUTATIONS, QUESTIONNAIRE VARIABLES

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
I1_i	Health during the past four weeks	21 hot deck	21	6,605	0.32
I9_i	Current health	34 hot deck	34	6,605	0.51
I17a_i	Wear glasses	15 hot deck	15	6,605	0.23
I17b_i	Difficulty seeing with glasses	9 logical, 28 hot deck	37	4,408	0.84
I18_i	Difficulty seeing no glasses	33 logical, 36 hot deck	69	2,234	3.09
I19_i	Uses special equipment because of difficulty seeing	49 logical, 12 hot deck	61	2,812	2.17
I21_i	Difficulty hearing	7 logical, 31 hot deck	38	6,605	0.58
I22_i	Able to hear normal conversation	26 logical, 23 hot deck	49	1,234	3.97
I23_i	Uses special equipment because of difficulty hearing	26 logical, 1 hot deck	27	1,234	2.19
I25_i	Difficulty having speech understood	4 logical, 36 hot deck	40	6,605	0.61
I26_i	Able to have speech understood at all	31 logical, 12 hot deck	43	1,691	2.54
I27_i	Uses special equipment because of difficulty speaking	31 logical, 2 hot deck	33	1,691	1.95
I29_i	Difficulty walking without assistance	14 logical, 43 hot deck	57	6,605	0.86
I30_i	Able to walk ¼ mile	24 logical, 51 hot deck	75	2,722	2.76

TABLE VII.4 *(continued)*

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
I31_i	Uses special equipment because of difficulty walking	24 logical, 8 hot deck	32	2,722	1.18
I33_i	Difficulty climbing 10 steps	8 logical, 69 hot deck	77	6,605	1.17
I34_i	Able to climb 10 steps at all	38 logical, 55 hot deck	93	2,852	3.26
I35_i	Difficulty lifting and carrying 10 lbs.	3 logical, 38 hot deck	41	6,605	0.62
I36_i	Able to lift or carry 10 lbs. at all	29 logical, 31 hot deck	60	2,537	2.36
I37_i	Difficulty using hands or fingers	21 hot deck	21	6,605	0.32
I38_i	Able to use hands or fingers at all	14 logical, 21 hot deck	26	1,498	1.74
I39_i	Difficulty reaching over head	40 hot deck	40	6,605	0.61
I40_i	Able to reach over head at all	28 logical, 17 hot deck	45	1,524	2.95
I41_i	Difficulty standing	59 hot deck	59	6,605	0.89
I42_i	Able to stand at all	31 logical, 9 hot deck	40	3,551	1.13
I43_i	Difficulty stooping	3 logical, 35 hot deck	38	6,605	0.58
I44_i	Able to stoop at all	14 logical, 42 hot deck	56	3,556	1.57
I45_i	Difficulty getting around inside home	2 logical, 19 hot deck	21	6,605	0.32
I46_i	Need help to get around inside home	17 logical, 3 hot deck	20	921	2.17
I47_i	Difficulty getting around inside home	3 logical, 30 hot deck	33	6,605	0.50
I48_i	Need help to get around outside home	16 logical, 21 hot deck	37	2,193	1.69
I49_i	Difficulty getting into/out of bed	4 logical, 28 hot deck	32	6,605	0.48
I50_i	Need help getting into/out of bed	20 logical, 11 hot deck	31	1,539	2.01

TABLE VII.4 *(continued)*

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
I51_i	Difficulty bathing or dressing	1 logical, 26 hot deck	27	6,605	0.41
I52_i	Need help bathing or dressing	17 logical, 9 hot deck	26	1,236	2.10
I53_i	Difficulty shopping	11 logical, 26 hot deck	37	6,605	0.56
I54_i	Need help shopping	16 logical, 10 hot deck	26	1,739	1.50
I55_i	Difficulty preparing own meals	7 logical, 24 hot deck	31	6,605	0.47
I56_i	Need help to prepare meals	17 logical, 9 hot deck	26	1,865	1.39
I57_i	Difficulty eating	20 hot deck	20	6,605	0.30
I58_i	Need help to eat	19 logical, 2 hot deck	21	685	3.07
I59_i	Trouble concentrating	42 hot deck	42	6,605	0.64
I60_i	Trouble coping with stress	80 hot deck	80	6,605	1.21
I61_i	Trouble getting along with people	70 hot deck	70	6,605	1.06
CageScore_indicator_i	CAGE Alcohol Score	30 constructed from imputed variables	30	6,605	0.45
I72_i	Use drugs in larger amounts than prescribed	51 hot deck	51	6,605	0.77

Source: NBS, round 3.

TABLE VII.5

HEALTH STATUS IMPUTATIONS, CONSTRUCTED VARIABLES

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
C_EquipFuncLim_I	Use equipment/device for functional/sensory limitation	20 constructed from imputed variables	20	6,605	0.30
C_NumSenLim_i	Number of sensory limitations	107 constructed from imputed variables	107	6,605	1.62
C_NumSevSenLim_i	Number of severe sensory limitations	52 constructed from imputed variables	52	6,605	0.79
C_NumPhyLim_i	Number of physical functional limitations	193 constructed from imputed variables	193	6,605	2.92
C_NumSevPhyLim_i	Number of severe physical functional limitations	191 constructed from imputed variables	191	6,605	2.89
C_NumEmotLim_i	Number of emotional/social limitations	149 constructed from imputed variables	149	6,605	2.26
C_NumADLs_i	Number of impaired activities of daily living (ADLs)	50 constructed from imputed variables	50	6,605	0.76
C_NumADLAssist_i	Number of ADLs requiring assistance	35 constructed from imputed variables	35	6,605	0.53
C_NumIADLs_i	Number of instrumental activities of daily living (IADL) difficulties	53 constructed from imputed variables	53	6,605	0.80
C_NumIADLAssist_i	Number of IADLs Requiring Assistance	35 constructed from imputed variables	35	6,605	0.53
C_PCS8TOT_i	Physical summary score	199 constructed from imputed variables	199	6,605	3.01
C_MCS8TOT_i	Mental summary score	199 constructed from imputed variables	199	6,605	3.01

TABLE IV.5 *(continued)*

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
C_DrugDep_i	Drug dependence	52 constructed from imputed variables	52	6,605	0.79

Source: NBS, round 3.

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

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

The only classing variable common to all imputations was the collapsed condition code variable. Age and gender were also used 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 presented in this section are constructed variables that pertain to nonemployment-based income. These include workers‘ compensation, private disability claims, unemployment, and other sources of regular income. The imputed variables in this section are described in Table VII.6.

TABLE VII.6

IMPUTATIONS ON SOURCES OF INCOME OTHER THAN EMPLOYMENT

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
C_AmtPrivDis_i	Amount received from private disability last month	77 logical, 13 imputed using specialized procedures	90	6,605	1.36
C_AmtWorkComp_i	Amount received from workers' compensation last month	37 logical, 8 imputed using specialized procedures	45	6,605	0.68
C_AmtVetBen_i	Amount received from veterans' benefits last month	29 logical, 10 imputed using specialized procedures	39	6,605	0.59
C_AmtPubAssis_i	Amount received from public assistance last month	38 logical, 16 imputed using specialized procedures	54	6,605	0.82
C_AmtUnemply_i	Amount received from unemployment benefits last month	30 logical, 1 imputed using specialized procedures	31	6,605	0.47
C_AmtPrivPen_i	Amount received from private pension last month	32 logical, 19 imputed using specialized procedures	51	6,605	0.77
C_AmtOthReg_i	Amount received from other regular sources last month	34 logical, 15 imputed using specialized procedures	49	6,605	0.74

Source: NBS, round 3.

In this section, respondents first were asked if they had received money from a specific source, and then the specific amount received from that source. If a respondent could not provide a specific value, he or she was asked a series of questions on whether the amount was above or below specific values, or was given the option of providing a range of values, where the options depended upon responses to a series of questions. After being classified according to a range of

values that he or she provided, the respondent was assigned the median of the specific values provided by 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 imputed first the range (using a random assignment), and then assigned the median of the values provided by respondents who gave specific values within that range. If the respondent did not know if he or she received funds from a source, we used hot-deck imputation to determine this, and then proceeded as above.

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

For variables requiring hot-deck imputation, the classing variables were the same for all variables: an indicator of whether the respondent was a recipient of SSI, SSDI, or both; living situation; and education. None of the variables requiring hot-deck imputation are listed in Table VII.6 because they were only component variables for the delivered variables listed in the table.

6. Section L: Personal and Household Characteristics

Other than the personal characteristics of race and ethnicity discussed earlier, most of the imputed variables in section L pertain to household characteristics. These questions include education (L3_i), marital status (L8_i), cohabitation status (C_Cohab_i), number of children in the household (C_NumChildHH_i), household size (C_Hhsize_i), and body mass index for poverty-level respondents (C_BMI_cat_i), since it is constructed of variables collected in section L. Most of these variables were imputed early in imputation processing and were used in the

imputation (FedPovertyLevel_cat1).[59] Also included in this section is the constructed variable for the work status variables; however, poverty level was imputed later. Both sets of variables are discussed in this section.

The imputation of poverty level required the imputation of annual income and household size. The annual income question was another case of a specific value being requested, and if a specific value could not be provided, the respondent was asked if the annual income fell within certain ranges. For this item, some respondents provided a specific value; some only 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 this table since it was not released in the final file. All of the missing values in C_FedPovertyLevel_cat1[60] were derived from the imputed annual incomes; hence all missing values are ~~–e~~onstructed from imputed variables." Table VII.7 identifies imputed variables in section L.

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

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

[60] 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 VII.7

IMPUTATIONS OF PERSONAL AND HOUSEHOLD CHARACTERISTICS

Variable Name	Description	Imputation Method	Number Missing	Number Eligible	Percent Imputed
C_BMI_Cat_i	Body Mass Index categories	246 hot deck	246	6,605	3.72
L3_i	Highest year/grade completed in school	87 hot deck	87	6,605	1.32
L8_i	Marital Status	32 hot deck	32	6,605	0.48
L11_i	Living arrangements	3 logical, 32 hot deck	35	6,605	0.53
C_NumChildhh_i	Number of children living in the household	26 hot deck, 4 constructed from imputed variables	30	6,605	0.45
C_hhsize_i	Household Size	35 hot deck	35	6,605	0.53
C_cohab_i	Cohabitation Status	3 logical, 34 hot deck	37	6,605	0.56
C_FedPovertyLevel_cat1	2005 Federal Poverty Level	2141 constructed from imputed variables	2141	6,605	32.41

Source: NBS, round 3.

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

VIII. USING THE NBS RESTRICTED AND PUBLIC USE FILES

A. FILE CONTENT AND TECHNICAL SPECIFICATIONS

The NBS round 3 Restricted Use File contain 6,605 records and 6,078 variables. The Public Use File contains 2,508 records and 704 variables. To reduce the risk of disclosure, individuals from the Ticket to Work participant sample are not included on the public file. Variables are positioned on the file in the following order:

Survey Administration Variables. These include variables related to survey administration including respondent type identifiers and other variables associated with the conduct of the survey.

Sampling Variables and Weights. These include administrative variables used for sampling purposes and administrative data that provide additional descriptive information about the sample.

Variables from Section A-M of the NBS Questionnaire. Variables are ordered within each section by related questionnaire item number. Constructed variables created from source variables within a section are ordered at the end of each section.

SSA Administrative Data. These include a select set of data from SSA administrative records to enhance analyses of Ticket to Work participants.

Both the Restricted Use File and Public Use File are available in a SAS ~~"sa~~s7bdat" format database. The Restricted Use File has the following technical specifications:

- Data set name: R3NBSRAF.
- Number of observations: 6,605.
- Number of variables: 6,078.
- Date last created: December 15, 2009.

The Public Use File has the following technical specifications:

- Data set name: R3NBSPUF.
- Number of observations: 6,605.
- Number of variables: 740.
- Date last created: December 15, 2009.

B. CHOOSING A SAMPLE AND WEIGHT VARIABLE

As discussed in Chapter II, the NBS comprises two independent samples: (1) the National Representative Beneficiary Sample and (2) the Ticket Participant Sample. Use of the appropriate weight variables allows estimates of either the national beneficiary population or the TTW participant population. The weights specified below should be used when performing any analysis. Due to the design of the NBS and the variation of weights within sampling strata, the use of unweighted rather than weighted data in the analysis will provide incorrect results.

Separate weights were computed for each sample to account for the sampling method, data collection method, and the target populations of the survey: one for the Representative Beneficiary Sample (R3_WTR3_ben), one for the Ticket Participant Sample (R3_WTR3cs_par), and one for the combined Representative Beneficiary and Ticket Participant sample (R3_WTR3_com). See Table VIII.1 for a summary of the appropriate weights for each population of interest.

TABLE VIII.1

NBS SAMPLE WEIGHTS

Weight Name	Description	Condition
R3_WTR3_Ben	Beneficiary weight	Orgsampinfo_Tstatus=2 (Representative Beneficiary Sample)
R3_WTR3CS_Par	Participant Weight	Orgsampinfo_Tstatus=1 (Ticket Participant Sample)
R3_WTR3_Com	Composite Weight	Orgsampinfo_Tstatus=1 or 2 (Combined Sample)

It is not necessary to subset the file when using the weights and, in fact, subsetting the file may result in incorrect estimates and problems with running the computer software.[61] The weights equal 0.0 for any case that is not in the analysis population. The variable OrgSampInfo_Tstatus identifies whether the case was selected for the Ticket Participant Sample (Orgsampinfo_Tstatus=1) or for the Representative Beneficiary Sample (Orgsampinfo_Tstatus=2). If the population of interest is the national beneficiary population, the Representative Beneficiary Sample weight (R3_WTR3_ben) should be used. This variable has a value greater than 0.0 for 2,508 cases (where Orgsampinfo_Tstatus=2) and a weight value of 0.0 for the 4,097 participant sample cases. If the population of interest is the TTW participant population, the Ticket Participant Sample weight (R3_WTR3CS_par) should be used. This variable has a value greater than 0.0 for 3,115 participant cases (where Orgsampinfo_Tstatus=1) and a weight value of 0.0 for the 3,490 beneficiary cases (where Orgsampinfo_Tstatus=2).

A composite sample weight (R3_WTR3_com) that combines the Ticket Participant Sample and the Representative Beneficiary Sample is also provided on the Restricted Use File (using all 6,605 cases). While this weight was provided to increase the sample size of the TTW participants for analyses of the national beneficiary population, it adds minimal additional analytic power. This weight can be used, however, in lieu of the Representative Beneficiary Sample weight (R3_WTR3_ben) for analysis of the national beneficiary population. When using the combined beneficiary and participant weight, the variable "flagparti" (rather than OrgSampInfo_Tstatus) can be used to identify Ticket participants (flagparti=1). As discussed in Chapter VI, 31 cases

[61] The design-based sampling variance estimate is best computed using the full data file because if subsetting is performed, some values for the design-based sampling variance parameters will be missing in the subset file. This can cause the software to provide incorrect sampling variance estimates or the computer program may fail to run properly.

sampled as part of the Representative Beneficiary Sample also appeared on the Ticket Participant sampling frame. These cases can be identified as participants when using the combined weight using the variable flagparti.

C. 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 are used to provide an essentially unbiased estimate of the sampling variance.

The NBS sample design involves stratification and unequal probabilities of selection. Variance estimates calculated from NBS data must incorporate the sample design features to obtain the correct estimate. Most procedures in atandard 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 is likely to substantially underestimate 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.[62]

The most appropriate sampling variance estimators for complex sample designs such as the NBS are the procedures based on the Taylor series linearization of the nonlinear estimator using explicit sampling variance equations, and the procedures based on forming pseudo-replications[63] of the sample. The Taylor series linearization procedure is based on a classic statistical method in which a nonlinear statistic can be approximated by a linear combination of the components within the statistic. The accuracy of the approximation is dependent on the sample size and the complexity of the statistic. For most commonly used nonlinear statistics (such as ratios, means, proportions, and regression coefficients), the linearized form has been developed and has good statistical properties. Once a linearized form of an estimate is developed, the explicit equations for linear estimates can be used to estimate the sampling variance. Because the explicit equations can be used, the sampling variance can be estimated using many features of the sampling design (for example, finite population corrections, stratification, multiple stages of selection, and unequal selection rates within strata). This is the basic variance estimation procedure used in SUDAAN, the survey procedures in SAS, STATA, and other software packages to accommodate

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

[63] Pseudo-replications of a specific survey sample, as opposed to true replications of the sampling design, entail the selection of multiple independent subsamples from the original sample data using the same sampling design. These subsamples can be random (as in a bootstrap) or restricted (as in Balanced Repeated Replication).

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, more survey data analysis software packages use the Taylor series linearization procedure and explicit sampling variance equations. Therefore, we developed the variance estimation specifications necessary for the Taylor series linearization (PseudoStrata and PseudoPSU). Example code for this procedure using SAS and the survey data analysis software SUDAAN is given in Appendix L.[64] Details about syntax for SAS are available from SAS (SAS Institute 2004). Details about SUDAAN syntax are available from RTI International (Research Triangle Institute 2004).

D. CODEBOOK

To aid the user, two codebooks were developed by MPR; one for the Restricted Use File and one for the Public Use Files. Both codebooks are available as separate reports and can be obtained from Mathematica Policy Research: "The National Beneficiary Survey: Round 3 Public Use File Codebook" (Wright et al. 2010) and "The National Beneficiary Survey: Round 3 Restricted Use File Codebook" (Wright et al. 2010).

The codebooks provide extensive documentation for each variable on the file including variable name, label, position, variable type and format, question universe, question text, number of cases eligible to receive each item, constructed variable specifications, and user notes. Frequency distributions and means are also included as appropriate.

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

IX. USING THE LONGITUDINAL DATA FILE

A. FILE CONTENT AND TECHNICAL SPECIFICATIONS

The longitudinal data file is comprised of Phase 1 Ticket participants who were part of the round 1 Ticket participant sample and completed round 1, 2 or 3, and Phase 2 Ticket participants who were part of the round 2 Ticket participant sample and completed round 2 or 3. Records are assigned a caseid which is the same across all completed rounds. The Restricted Use File contains 17,841 variables and 1,728 observations. There is no longitudinal Public Use File. Variables are positioned in the same order they appear on the cross-sectional files, with round 1 variables first, followed by round 2, and round 3 variables. All variables have a prefix (R1_, R2_, or R3_) to indicate which survey round and instrument they are from.

B. CHOOSING A SAMPLE AND WEIGHT VARIABLE

There are two longitudinal samples in the National Beneficiary Survey:

1. ***The Phase 1 Longitudinal Sample.*** These cases are TTW participants selected in round 1. This sample consists of 1,466 participants, all residing in Phase 1 states. All Phase 1 longitudinal cases, with the exception of those who died or were incarcerated after round 1 sample selection, were fielded at rounds 2 and 3, regardless of their completion status at round 1.
2. ***The Phase 2 Longitudinal Sample.*** These cases are TTW participants selected in round 2 who resided in Phase 2 states. This sample consists of 1,350 participants. All Phase 2 cases were fielded at round 3, with the exception of those who died or were incarcerated after round 2 sample selection, regardless of their completion status at round 2.

Table IX.1 shows the completion status of Phase 1 and Phase 2 longitudinal cases across all three rounds.

TABLE IX.1

LONGITUDINAL TICKET PARTICIPANT SAMPLE COMPLETION STATUS ACROSS ROUNDS

	Completed R1 & R2		Completed R2 & R3		Completed R1 & R3		Completed R1, R2, & R3		
Sample	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Total Sample
Phase 1	897	61.2	845	57.6	837	57.1	767	52.3	1466
Phase 2			831	61.6					1350

Source: NBS.

A non-response bias analysis demonstrated that there was no significant bias in comparisons between the following groups: (1) all round 1 sample cases; (2) round 1 sample cases who responded in round 3; and (3) round 1 sample cases that responded in all 3 rounds on key variables. Hence, a decision was made to focus analyses on Phase 1 cases that responded in all three rounds. Weights were also created for Phase 1 cases that completed rounds 1 and 2 and Phase 2 cases that completed rounds 2 and 3 to permit an examination of Phase 1 and Phase 2 individual-level changes over one year for two yearly cohorts of TTW participants. Therefore, the following weights are included on the longitudinal data file:

1. ***PARTFNLWGT_LONG123***. This weight was constructed for Phase 1 TTW participants who completed all three rounds of data collection (round 1, round 2, and round 3). These cases (n=767) can be identified using the variable R3_STATUSR1R2R3=3 (Longitudinal, completed R2 and R3/completed R1, R2, and R3). These cases are also coded as R3_R1LONG=1 and R3_ORGSAMPINFO_PHASE=1 which identify phase 1 longitudinal cases on the Round 3 file.[65]
2. ***PARTFNLWGT_LONG12***. This weight was constructed for Phase 1 TTW participants who completed rounds 1 and 2 of data collection, regardless of their

[65] There are 11 cases with positive weights for PARTFNLWGT_LONG123, but with values of R3_STATUSR1R2R3 and R3_R1LONG both missing. This is due to the fact that sample members who were ineligible when surveyed were treated as respondents for the purposes of calculating weights. These 11 individuals were either deceased or were incarcerated at the time of the Round 3 survey. They were included on the file because they were among the 897 Phase 1 cases with complete survey information for Rounds 1 and 2.

round 3 status. These cases (n=897) can be identified using the variable R2_STATUSR1R2=2 (Longitudinal Participant Completed R1 and R2). These cases are also coded as R2_R1LONG=1 and R2_ORGSAMPINFO_PHASE=1 which identify phase 1 longitudinal cases on the Round 2 file.

3. ***PARTFNLWGT_ LONG23***. This weight was constructed for Phase 2 TTW participants who completed all rounds for which they were fielded (Rounds 2 and 3). These cases (n=831) can be identified using the variable R2_STATUSR1R2=3 (New Participant/Longitudinal Participant Did Not Complete R1, Completed R2) or R3_STATUSR1R2R3=2 (Longitudinal, completed R2 and R3 (Phase 2)). These cases are also coded as R3_R2LONG=1 and R2_ORGSAMPINFO_PHASE=2 which identify phase 2 longitudinal cases on the Round 3 file.

C. CODEBOOK

To aid the user, a codebook was developed by MPR for use with the Restricted Use Longitudinal File. The codebooks is available as a separate reports and can be obtained from Mathematica Policy Research: ―The National Beneficiary Survey: Longitudinal Ticket Participant Codebook" (Wright et al. 2008).

The codebooks provide extensive documentation for each variable on the file including variable name, label, position, variable type and format, question universe, question text, number of cases eligible to receive each item, constructed variable specifications, and user notes. Frequency distributions and means are also included as appropriate.

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

REFERENCES

Bethel, J. and D. Stapleton. "Evaluation Design For The Ticket To Work Program: Final Survey Sample Design." Prepared for the Social Security Administration. Washington, DC 2002.

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

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

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

Grau, E., D. Wright, Y. Zheng, and Potter, F. "National Beneficiary Survey Round 3 (Volume 1 of 3): Editing, Coding, Imputation, and Weighting Procedures." Report submitted to the Social Security Administration. Washington, DC: Mathematica Policy Research, Inc., December 2008.

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.

Mitchell, S., A. Ciemnecki, K. Cybluski, and J. Markesich, J. "Removing Barriers to Survey Participation for Persons with Disabilities." Prepared for Cornell University. Washington DC: The Employment and Disability Institute, 2004.

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.

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

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

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

U.S. Department of Health and Human Services, Health Resources and Services Administration, Bureau of Health Professions. Area Resource File (ARF). Rockville, MD: USDHHS, February 2003.

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

Wright, D., Y. Fogus, and S. Khambhati. "The National Beneficiary Survey: Longitudinal Ticket Participant Codebook" Prepared for the Social Security Administration. Washington DC: Mathematica Policy Research, Inc., 2008.

Wright, D., S. Khambhati, and B. Jones. "The National Beneficiary Survey: Round 3 Public Use File Codebook" Prepared for the Social Security Administration. Washington, DC: Mathematica Policy Research, Inc., 2010.

Wright, D., S. Khambhati, and B. Jones. "The National Beneficiary Survey: Round 3 Restricted Use File Codebook" Prepared for the Social Security Administration. Washington DC: Mathematica Policy Research, Inc., 2010.

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