MATHEMATICA
Policy Research, Inc.

National Beneficiary Survey Round 1 (Volume 2 of 3): Data Cleaning and Identification of Data Problems

December 19, 2008

Debra Wright
Kirsten Barrett

ERRATA

(Updated December 20, 2016)

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

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

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

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

CONTENTS

CONTENTS *(continued)*

Chapter   **Page**

CONTENTS *(continued)*

TABLES

TABLES *(continued)*

ACRONYMS

ADLs: Activities of Daily Living

CAPI: Computer-assisted personal interviewing

CATI: Computer-assisted telephone interviewing

ENs: Employment Networks

IADLs: Instrumental Activities of Daily Living

ICD-9:   International Classification of Diseases - 9th revision

IWP: Individual Work Plan

MPR: Mathematica Policy Research

NAICS: North American Industry Classification System

NBS: National Beneficiary Survey

SOC: Standard Occupational Classification

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)

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

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

A.   NBS SAMPLE DESIGN OVERVIEW

SSA implemented the TTW program in three phases spanning three years, with each phase corresponding to about one-third of the states. The initial NBS survey design called for four

---

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

[2] The Ticket to Work program, implemented in 2002, was phased in nationwide over three years. In 2002, the

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

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

---

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

TABLE I.1

NATIONAL BENEFICIARY AND TTW PARTICIPANT SAMPLE SIZES

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

Source: 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.

TABLE I.2

ROUND 1 SAMPLE SIZES, TARGET COMPLETES, AND ACTUAL COMPLETES

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

Source: NBS, round 1

For all survey rounds, the NBS used a multi-stage sampling design with a supplemental single-stage sample for some Ticket participant populations. For the multi-stage design, data from SSA on the counts of eligible beneficiaries in each county were used to form primary sampling units (PSUs), consisting of one or more counties. The sample of all SSA beneficiaries

(the Representative Beneficiary Sample) was selected from among beneficiaries residing in these PSUs (or, in the case of two counties with a large number of beneficiaries, from secondary sampling units) using age-defined sampling strata. Separate samples of Ticket participants within each phase in the original sample design were selected from all Ticket participants in these PSUs. The Ticket Participant Sample was divided into three strata within each phase according to the types of payment system under which SSA paid a service provider: (1) the traditional vocational rehabilitation payment system, (2) the milestone-outcome payment system, or (3) the outcome-only payment system.[3] The supplemental single-stage sample for some Ticket participant populations was drawn from all Ticket participants, not just those in the PSUs, with stratification based upon payment type and whether the participant was in a PSU or not. The Round 1 User's Guide (Wright et al. 2008) contains more information on the round 1 sampling design.

B. NBS OBJECTIVES

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

The NBS has five key objectives:

---

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

1. To provide critical data on the work-related activities of Supplemental Security Income (SSI) and Social Security Disability Insurance (SSDI) beneficiaries, particularly as they relate to the TTW implementation

2. To collect data on the characteristics and program experiences of beneficiaries who use their Ticket

3. To gather information about beneficiaries who do not use their Tickets, 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

The round 1 NBS survey data will be combined with SSA administrative data to provide critical information on access to jobs and employment outcomes for disability beneficiaries, including those who participate in the TTW program and those who do not. Though some sections of the NBS target beneficiary activity directly related to TTW, most of the survey captures more general information on SSA beneficiaries including their disabilities, interest in work, use of services, and employment. As a result, SSA and external researchers interested in disability and employment issues can use the survey data for other policymaking and program-planning efforts.

C. ROUND 1 DATA COLLECTION OVERVIEW

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

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

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

The round 1 CATI and CAPI data collection was completed in October 2004. Interviews were completed with 6,520 individuals in the Representative Beneficiary Sample and 1,083 people in the Ticket Participant Sample for a total of 7,603 interviews completed. An additional 458 beneficiaries and 73 Ticket participants were determined to be ineligible to participate in the survey.[4] Across both samples, 6,302 cases were completed by telephone and 1,301 were completed by CAPI. Proxy interviews were completed for 1,997 sample members. There were 109 cases in which the sample member was unable to participate and a proxy was unable to be identified. A total of 184 Spanish-language interviews were completed. The weighted response rate for the Representative Beneficiary Sample was 77.5 percent. The weighted response rates for the Ticket Participant Sample was 80.9 percent. More information about the sample selection and sampling weights can be found in Potter et al. (2008).

---

[4]Ineligible sample members include those who were deceased, no longer living in the continental United States, incarcerated, and those whose benefit status was pending. For Ticket participants, ineligibles also included sample members who left the program after sampling was completed, or were not contacted by telephone and were not eligible for field follow-up.

D. NBS DATA DOCUMENTATION REPORTS

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

- ***Editing, Coding, Imputation, and Weighting Report*** (Potter et al. 2008). This report summarizes the editing, coding, imputation, and weighting procedures as well as the development of standard errors for the round 1 NBS. It includes an overview of the variable naming, coding, and construction conventions used in the data files and accompanying codebooks; describes how the 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*** (current report). This report describes the data processing procedures performed for round 1 of the NBS. It outlines the data coding and cleaning procedures and describes the data problems identified, their origins, and the corrections implemented to create the final data file. The report describes the data issues by sections of the interview and concludes with a summary of types of problems encountered and general recommendations.

- ***User's Guide for Restricted and Public Use Data Files*** (Wright et al. 2008). This report is designed to provide users with information about the restricted use data file and planned public use file, including construction of the files; weight specification and variance estimation; masking procedures employed in the creation of the Public Use File; and a detailed overview of the questionnaire design, sampling, and NBS data collection. The report also contains information covered in the two reports mentioned above including procedures for data editing, coding of open-ended responses, and variable construction; and a description of the imputation and weighting procedures and development of standard errors for the survey.

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

- *NBS Questionnaire*. This document contains all items on the round 1 survey and includes documentation of skip patterns, question universe specifications, text fills, interviewer directives, and consistency and range checks.

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

In the discussion that follows, we describe the data processing procedures MPR performed for round 1 of the NBS. An extensive review of the NBS data was conducted in order to identify data problems prior to analysis. This report 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. We begin with a brief overview of the NBS instrument. We then describe data issues by sections of the interview and conclude with a summary of types of problems encountered and general recommendations.

II. DESCRIPTION OF THE NBS INSTRUMENT

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 pre-testing. More information about the questionnaire can be found in the Round 1 User's Guide (Wright et al. 2008). The survey instrument is available from MPR upon request (Mitchell et al. 2004).

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.[5] Measurements were treated in a similar way. Thus, questions that mentioned a specific weight also mentioned the kilogram equivalent.[6] Interpreters were used as needed to conduct interviews in languages other than Spanish.

A. SUMMARY OF MODULES

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

---

[5] 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?

[6] 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 A—Introduction and Screener

- Section B—Disability and Current Work Status

- Section C—Current Employment

- Section D—Jobs/Other Jobs During 2003

- Section E—Awareness of SSA Work Incentive Programs and Ticket to Work

- Section F—Ticket Non-Participants in 2003

- Section G—Employment-Related Services and Supports Used in 2003

- Section H—Ticket Participants in 2003

- 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—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, those received, as well as those needed but not received. Other questions solicit information about job satisfaction.

4. Section D—Jobs/Other Jobs During 2003

This section collects information about employment during the 2003 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 Ticket to Work program, information is collected on how beneficiaries learned about the program and the names and dates they signed up with their current service providers.

6. Section F—Ticket Non-Participants in 2003

This section is administered to beneficiaries not participating in the TTW program and collects data on reasons for non-participation. 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 2003

Questions in this section ask beneficiaries about their use of employment-related services in calendar year 2003, 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 2003

This section asks 2003 Ticket 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 TTW 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[TM][7] scale), difficulties with Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs), a variety of functional limitations, substance abuse/dependence, and treatment for mental health conditions.

10. Section J—Health Insurance

Questions in this section collect information about the sources of health insurance coverage the beneficiary has, both at the time of interview and during calendar year 2003.

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.

---

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

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 PRELOADED DATA

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 Ticket). Similarly, both CATI and CAPI respondents received the same questionnaire. The NBS took, on average, 45 minutes to administer. The interview length ranged from 15 to 180 minutes excluding TTY, TRS, and instant messaging interviews.

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 2003 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 2003. Only respondents who reported using their Ticket to sign up with a provider in

2003 were asked questions from section H. See Table II.1 for a summary description of the main

questionnaire pathing.

TABLE II.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 2003	Respondents who answer (B30 = YES) Question B30: Did you work at a job or business for pay or profit anytime in 2004?
E	Awareness of SSA Work Incentive Programs and Ticket to Work	All respondents
F	Ticket Non-Participants in 2003	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 2003?
G	Employment-Related Services and Supports Used in 2003	All respondents
H	Ticket Participants in 2003	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 2003?
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 1.

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 II.2 for a complete list and description of preloaded variables.

TABLE II.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.

16

TABLE II.2 *(continued)*

Variable	Definition	Purpose
Phase	Phase of TTW roll-out. Based on the sample member's state of residence at the time the sample frame was drawn.[a]	Used to determine pathing at item B32 (worked for pay before November 1, 2003). Only respondents who worked for pay in 2003 and who resided in a Phase 3 roll-out state were asked this item.
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 1.

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

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 II.3 for a complete list of items that were not asked of proxy respondents.) Proxy interviews were completed for 1,999 cases.

TABLE II.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:
H10a	Participation in the Ticket to Work program is voluntary and you do not have to participate to keep your disability benefits.
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 1.

III. ROUND 1 DATA PROCESSING

A. CODING OF OPEN-ENDED AND VERBATIM RESPONSES

The NBS questionnaire includes a number of questions designed to elicit open-ended responses. To make it easier to use the data connected with these responses in an analysis, we grouped the responses and assigned them numeric codes when possible. The methodology used to code each variable depended upon the content of the variable. 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 pre-defined response category without reading the categories aloud to the respondent. If none of the response options seem to apply, interviewers select an "other specify" category and type in the response.

As part of data processing, we examined a portion of all verbatim responses in an attempt to uncover dominant themes for each question. Based on this initial review, we developed a list of categories and decision rules for coding verbatim responses to open-ended items. In addition, supplemental response categories were added to some field-coded or other/specify items to facilitate coding if there were enough such responses and they could not be back-coded into pre-

existing categories. Thus we categorized verbatim responses for quantitative analyses by coding responses that clustered together (for open-ended and "other/specify" responses) or by back-coding responses into existing response options if appropriate (for "field-coded" and "other/specify" items). If during coding, it became apparent that changes to the coding scheme were necessary (for example adding additional categories or clarifying coding decisions), new decision rules were discussed and documented. Verbatim responses were sorted alphabetically by item for coders and could be filtered by coding status so that new decision rules could be easily applied to cases that had been previously coded. When it was impossible to code a response, when responses were invalid, or when they could not be coded into a given category, we assigned a two-digit supplemental code to the response (see Table III.1). The verbatim responses themselves are excluded from the data files. Chapter IV indicates which items in each instrument section required coding and lists any additional response categories created during coding.

TABLE III.1

SUPPLEMENTAL CODES FOR OTHER/ SPECIFY CODING

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

Source: NBS, round 1.

B. DATA CLEANING

Once coded data were incorporated, a preliminary data file was created. A systematic review of the frequency counts of the individual questionnaire items was then conducted. Frequency counts were reviewed by each questionnaire path to identify possible skip pattern errors. Additionally, interviewer notes and comments were reviewed as a means to flag and correct individual cases. In consultation with SSA and research analysts, we took the general approach of editing only those cases where there appeared to be an obvious data entry or respondent error. As a result, while a substantial amount of time was spent meticulously reviewing individual responses, some suspect values remain on the file.

For all items with fixed field numeric responses (such as number of weeks, number of jobs, dollar amounts, and so on) we reviewed the upper and lower values that had been assigned by interviewers. While data entry ranges were set in the CATI instrument to prevent the entry of improbable responses, these ranges were intentionally set to encompass a wide range of values to account for the diversity expected in this population, and so that the interview could continue in most situations. Extremely high and low values were set to missing if it appeared that a data entry error had been made. Several consistency edit checks were also included throughout the NBS instrument to flag potential problems during the course of the interview. To minimize respondent burden, however, all consistency edit checks were suppressible. While the interviewer was instructed to probe such responses, the interview could continue past the item if the respondent could not resolve the problem. In the post-interview stage, we manually reviewed consistency problems that remained to determine if the responses were plausible. These cases were investigated and corrected or set to missing when an obvious error was encountered.

During data processing, we created several constructed variables to combine data across items. For these items, both the survey team and the analysis team reviewed the specifications,

21

multiple reviewers checked the SAS programming code, and we reviewed all data values for the constructed variables based on the composite variable responses and frequencies.

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

C. IDENTIFICATION OF DATA PROBLEMS

The data problems we identified in the course of checking the data file can be characterized as either measurement error or processing error. Measurement error is the difference between the observed value of a variable and the true, but unobserved, value of that variable. Sources of measurement error can include the questionnaire itself (including design, format, and content), the data collection mode, the interviewer, and the respondent. As we will discuss below, it is likely that the questionnaire, interviewer, and respondent all contributed to the data problems in the NBS. Processing errors as discussed in this report consist of incorrect specification or implementation of a complicated skip pattern or edit. In this report we discuss programming errors that resulted in incorrect skip patterns in the NBS. This report focuses on the identification of measurement errors and processing errors at the individual item level.

The identification of data problems on the NBS file occurred at several steps during the data cleaning and data preparation process. Many errors were identified through the systematic review of the frequency counts of the individual questionnaire items. Other data problems were identified during the development of the constructed variables and the implementation of the imputation procedures. Chapter IV describes the results of the review by instrument section. Recommendations to improve the quality of future data files are also made.

Despite instrument complexities, programming errors resulted in a relatively small number of issues related to data quality, as illustrated throughout this report. This is due, in large part, to the MPR staff's extensive internal testing of the NBS instrument before fielding the survey. In addition, as a matter or practice, we reviewed frequencies for the first 100 completed cases in round 1. This provided an additional opportunity to identify any programming issues that may have been missed during the internal testing. However, problems were identified as a result of programming, specification, or interviewer errors. A summary of the main problems encountered is presented in Table III.2. These issues are described in more detail in Chapter IV.

TABLE III.2

MAIN PROBLEMS ENCOUNTERED

Item	Description of Problem
B22 (Working for pay when first became limited)	Due to a programming error, age of limitation was not calculated for those who reported the year of their limitation rather than their age (B18_age=99). This resulted in 130 cases incorrectly skipping item B22. These cases are coded as .M (indicating missing due to error) in the data file. This problem is discussed in more detail in Chapter IV, Section B.2.
E3-E10 (Awareness and use of PASS, Earned Income Exclusion, PESS, and Continued Medicaid Eligibility)	All SSI beneficiaries were to be asked questions E3-E10. Due to a specification error, beneficiaries who received both SSI and SSDI incorrectly skipped these questions. A second problem occurred when respondents were unable to complete the interview in one session. When the interview continued in a second session, the information about the sample member's SSI and DI status was not saved in the system, resulting in these cases incorrectly skipping these items. These cases are coded as .M in the data file (2,762 cases). A special set of weights also was calculated to account for the missing data. This problem is discussed in more detail in Chapter IV, Section E.1.
E12-E13 (Awareness and use of Student Earned-Income Exclusion)	SSI beneficiaries who were 25 or younger at time of sampling and who received SSI benefits before age 22 were to be asked questions E12-E13. Due to the problems mentioned above at E3-E10, beneficiaries who received both SSI and SSDI, or who did not complete these items in the first session, incorrectly skipped these questions. In addition, information about the sample member's SSI and DI status and age at interview was not saved in the system. These cases are coded as .M in the data file (579 cases). A special set of weights also was calculated to account for the missing data. This problem is discussed in more detail in Chapter IV, Section E.1.

TABLE III.2 *(continued)*

Item	Description of Problem
E15-E18 (Awareness and use of Trial Work Period and Extended Period of Eligibility for Medicare)	All DI beneficiaries were to be asked questions E15-E18. Due to the problems mentioned above at E3-E10 and E12-E13, beneficiaries who received both SSI and SSDI, or who did not complete these items in the first session, incorrectly skipped these questions. In addition, information about the sample member's SSI and DI status was not saved in the system. These cases are coded as .M in the data file (2,728 cases). A special set of weights also was calculated to account for the missing data. This problem is discussed in more detail in Chapter IV, Section E.1.
E24 (Awareness of Ticket to Work participation)	Ticket to Work participants who said that they had not heard of the Ticket to Work program in items E21 and E22 were to be asked E24. Because the information about the sample member's Ticket participation status was not saved in the system when the interview continued in a second session, Ticket participants who did not complete this item in the first interview session incorrectly skipped it. In the data file, cases that incorrectly skipped these items are coded as .M (45 cases). Because E24 was used only to prompt Ticket to Work participants to recall the name of the program, revised weights were not calculated for this item. This problem is discussed in more detail in Chapter IV, Section E.1.
G13-G14 (Type of provider supplying job training)	To aid in the recall of employment-related services received in 2003, 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 then was collected for each provider mentioned. To minimize respondent burden by avoiding the need to ask provider type again, interviewers could indicate that a provider already had been mentioned at G12, thus skipping the provider type follow-up questions. In some cases, however, interviewers indicated that a provider already had been mentioned, when it had not. This resulted in missing data on the provider type questions. In these cases, G13 and G14 were coded as .M (23 cases). This problem is discussed in more detail in Chapter IV, Section G.1.
G18 (Type of provider supplying medical services)	As for G13 and G14, in some cases, interviewers incorrectly indicated at G17 that a provider already had been mentioned when it had not. For this reason, provider type is missing for 97 cases at item G18. These cases are coded as .M in the data file. This problem is discussed in more detail in Chapter IV, Section G.1.
G22 (Type of provider supplying mental health services)	As for G13, and G14, interviewers incorrectly indicated at G21 that a provider already had been mentioned when it had not. For this reason, G22 is missing for 104 cases. These cases are coded as .M in the data file. This problem is discussed in more detail in Chapter IV, Section G.1.

TABLE III.2 *(continued)*

Item	Description of Problem
G33 and G34 (When last received services from providers)	Once a list of providers was obtained, respondents were asked when they last had 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 2003 were asked about each provider from whom services were received in 2003. Before asking when services were received, the list of providers given at items G2, G11, G16, and G20 was compiled for the interviewer, who was asked to verify if any of the providers on the list were duplicates. The interviewer also was asked to verify with the respondent if any of the providers on the list were the same. Providers marked as duplicates were removed from the list and the provider-specific follow-up information was not obtained. There were some cases in which providers were marked as duplicates, but did not appear to be duplicates based on an examination of provider name and type. These cases were coded as .M in items G33 and G34, since the followups regarding when services were received were not asked of these providers. These cases are coded as .L in G36-G47_year_34. Across all providers, there are 456 instances in which this data is missing. This problem is discussed in more detail in Chapter IV, Section G.1.
K3 (Earnings last month before taxes and deductions)	To avoid asking K3 of respondents who already had indicated in Section B that they had never worked, or who had indicated in Section C that they had not worked in the month before the interview, a series of checks were programmed to skip this item for respondents not working last month. However, these checks were based on B24 (currently working) and B36 (ever worked) only. They did not take into account respondents who indicated in B30 (worked in 2003) or in B22 (worked when limited) that they had worked. These respondents should have been asked item K2A, "Did you work last month?" but instead inappropriately skipped to items K3 (how much earned last month), and K3a (how much was left after taxes and deductions). This programming error affected 3,274 cases. Cases with this problem are coded as .M (missing due to error) in item K2a. They are coded as .L (logical skip) in K3 and K4. This problem is discussed in more detail in Chapter IV, Section K.1.

Source: NBS, round 1.

IV. SECTION-BY-SECTION DESCRIPTION OF FINDINGS

A. SECTION A—SCREENER

The NBS screener was designed to identify and gain the cooperation of the respondent in addition to verifying that the sample person was still eligible for the survey. It was also used to determine if the sample member was capable of completing the interview and if the sample member required special accommodations such as TTY, TRS, or an in-person interview.

1. Date of Birth

Sample member name and date of birth from SSA records were used to verify that the correct person has been contacted. If two of the three date of birth elements provided by SSA matched the self-reported information (for example month and year match), the interview continued. If one or fewer elements matched, the interview was terminated and the case sent to locating. If the respondent could not provide a date of birth, the age of the sample member was requested. If the age was with two years (plus or minus), the interview continued.

Of the successfully screened respondents, there were 59 cases in which the date of birth collected was different from the date of birth provided by SSA. In 54 percent of these cases, the year of birth was off by one year. In 34 percent of cases, the year was off by two to nine years. These discrepancies were not edited and remain on the file. For cases that differed by 10 or more years (seven cases), the year of birth was set to equal the year of birth from SSA records since these appeared to be the result of data entry errors. Additionally, for cases in which age was provided in lieu of date of birth (six cases), the date of birth from SSA records was filled in for self-report date of birth (A68, A68a, and A68b).

27

2. Discrepancies in Respondent Type

Three screener items were used to determine if the sample member was cognitively able to participate in the survey process. These items addressed key elements of informed consent—the study topics, the voluntary nature of participation, and confidentiality. If the sample member did not pass any of the three items (within two attempts), a proxy respondent was sought. In order for the proxy to complete the survey on the sample member's behalf, the proxy was also required to pass the cognitive screener. Additionally, interviewers could complete the interview with a proxy if a knowledgeable informant indicated that the sample person would not be able to participate even with an accommodation, or if it became clear during the course of the interview that the sample person was not capable of responding. Interviewing 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.

At the end of Section A, the interviewer was asked to indicate whether the respondent to the survey was a sample member or proxy. This information was used to create the constructed variable, C_Rtype (Respondent Type). At the end of an NBS interview, the interviewer recorded whether the sample member or proxy completed the majority of the survey. In most cases, these two items were congruent. However, in 214 cases they were discrepant. That is, a sample member began the interview and a proxy completed the majority of it or vice-versa. Switching respondents was anticipated. It was expected that a small number of sample members would pass the cognitive screener but would be unable to recall or report information for the vast majority of questions in the survey. Cases where there were discrepancies were reviewed to determine if an interview error was made in coding the respondent. In general, we considered the interviewer data collected at the time the survey was completed as the most accurate for the purposes of

28

creating this construct. That is, if the sample member began the interview but the interviewer indicated that the proxy completed most of it, respondent type was recoded to proxy.

In terms of the survey questions, perception and attitudinal questions were asked only of sample members. These questions focus on overall job satisfaction and satisfaction with various work characteristics, awareness of and satisfaction with the TTW program, and satisfaction with state vocational rehabilitation services. Specific items skipped for proxies include C18, C39, H10, H11, H45, H58, and H59. In round 1, there are cases in which the sample member answered some of the sample member-only items and then a proxy stepped in and completed the rest of the survey. As described above, in this case, respondent type was recoded as proxy. At round 1, data from sample members on these early items were retained. For future rounds, we recommend filtering out such responses during the post-editing process to avoid confusion in following instrument pathing and analysis.

B. SECTION B—DISABILITY AND WORK STATUS

Section B contained questions on the sample member's limiting physical or mental condition(s) and employment status. This section also included questions designed to determine what job characteristics were important to sample members and collected information about work-related goals and expectations.

1. Health Condition Coding

In Section B of the questionnaire, each respondent was asked to cite the main and secondary physical or mental conditions that limit the kind or amount of work or daily activities they can do. Main conditions could be reported at one of four items: B2 (main reason limited), B6 (main reason eligible for benefits), B12 (main reason formally eligible for benefits if not currently eligible), and B15 (main reason limited when first started getting disability benefits). The

29

majority of respondents (92 percent) reported a main limiting condition at B2. The main purpose of items B6, B12, and B15 was to collect information on a health condition from people who reported no limiting conditions in B2. For example, if respondents said that they had no limiting conditions, they were asked if they were currently receiving benefits from Social Security. If they answered "yes," they were asked for the main reason that made them eligible for benefits (B6). If respondents said that they were not currently receiving benefits, they were asked whether they had received disability benefits in the last five years. If they answered "yes," they were asked for the condition that made them eligible for Social Security benefits (B12) or for the reason that first made them eligible if they no longer had that condition (B15). If respondents said that they had not received disability benefits in the last five years, they were screened out of the survey and coded as ineligible. Each response to B2, B6, B12, and B15 was assigned a value for the three constructs. Although respondents were asked to cite one "main" condition in B2, B6, B12, or B15, many listed more than one. These additional responses were maintained under the main condition variable and coded in the order in which they were recorded.

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

30

asked for the main reason they were first limited were also asked if there were any other conditions that limited them when they first started receiving benefits (B17).

The respondents' verbatim responses were coded using the International Classification of Diseases, 9th revision, Clinical Modification (ICD-9-CM) five-digit coding scheme. The ICD-9 is a classification of morbidity and mortality information developed in 1950 to index hospital records by disease for data storage and retrieval. The ICD-9 was available in hard copy for each 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). More information on coding responses to the health condition items can be found in "The National Beneficiary Survey: Round 1 Editing, Coding, Imputing, and Weighting Procedures" report (Potter et al. 2008).

Following ICD-9 coding, a series of constructed variables were created to group the health conditions reported at B1 and B2 into four different classifications of broad disease groups. 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.

a. Multiple Main Conditions

Health condition coding of respondent-provided data is complex. Often respondents do not know the name of the condition, or describe it in vague terms (for example, "he is slow" or "she has trouble breathing"). Although respondents were asked to provide one "main" condition in B2, B6, B12, or B15, many listed more than one. Rather than attempt to discern which listed condition was the main condition when more than one was given, conditions were coded in the

31

order provided by the respondent and named on the file as _1, _2, and so on. We recommend for future rounds that the importance of collecting a main condition be emphasized in interviewer training to limit the number of conditions reported for this item.

b. Duplicate Conditions

In approximately 260 cases (2 percent of cases) respondents mentioned a condition twice when reporting their main condition or reported a secondary condition that had already been reported as a main condition. During the process of coding such responses, coders identified any duplicate conditions by assigning the code 96. Additionally, during data cleaning and editing, ICD-9 codes within and across main and secondary items were compared to check for duplicate codes. Duplicates that were identified during coding or cleaning that followed valid codes were then dropped. In the event that the only condition reported was a duplicate of the main condition, the code was dropped and the filter item ("Do you have any other physical or mental conditions that limit the kind or amount of work or other daily activities you can do?") was recoded to "no."

c. Uncodeable Conditions

We anticipated that not all verbatim responses would contain enough information to allow coders to assign a specific ICD-9 code. To handle these situations, we provided coders with supplemental two-digit codes, which mirrored the chapter-level headings in the ICD-9 index, to allow a general code to be assigned in these instances (see Table IV.1). Approximately 3 percent of the verbatim responses coded at each medical condition item could not be coded to a specific ICD-9 code and were assigned a two-digit supplemental code.

TABLE IV.1

ICD-9 CATEGORY AND SUPPLEMENTAL CODES

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

TABLE IV.1 *(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.	No ICD-9 codes
95	Refused	Verbatim indicates respondent refused to answer the question.	No ICD-9 codes
96	Duplicate condition reported	The condition has already been coded for the respondent.	No ICD-9 codes
97	No condition reported	The verbatim does not contain or symptom to condition to code.	No ICD-9 codes
98	Don't know	The respondent reports that he/she does not know the condition.	No ICD-9 codes
99	Uncodeable	A code cannot be assigned based on the verbatim response.	No ICD-9 codes

Source: NBS, round 1.

In addition, in 111 cases (one percent) the verbatim responses could not be coded into either an ICD-9 code or broader two-digit supplemental code. In these cases, responses were coded by coders as "don't know" (code 98), "refused" (code 95), "uncodeable" (code 99), or "no condition reported" (code 97). Although this happened relatively infrequently in round 1, we recommend in future rounds increasing the amount of time spent training interviewers to collect disability and medical condition information from respondents. This would involve providing interviewers with more guidance on obtaining codeable information, using examples of uncodeable responses from round 1 to illustrate.

2. Working for Pay When First Limited

Item B18_age asks at what age the sample person first became limited. Respondents who could not provide an age were asked to give the year they were first limited (B18_year). Item B22 (working for pay when first became limited) was to be asked of all sample members whose age of limitation was 18 or older. Due to a programming error, age of limitation was not calculated for those who reported the year of their limitation rather than their age (B18_age=99). This resulted in some cases where respondents said they were limited after the age of 18 incorrectly skipping item B22 (130 cases). These cases are coded as .M (missing in error) on the data file. In future rounds, this will be corrected so that age of limitation is calculated for all cases within the instrument prior to B22.

3. Goals Include Not Receiving SSA Benefits

Item B37 asked if the sample person's personal goals included getting a job, moving up in a job, or learning new job skills. In round 1, only respondents who answered "yes" to B37 were asked B37a ("Do your personal goals include someday working and earning enough to stop receiving Social Security disability benefits?"). It was later determined that this item should be

asked of all respondents, not just those answering "yes" to B37. In subsequent rounds, all respondents should receive both B37 and B37a.

4. Back-Coding Responses to "Other/Specify" Items

Item B25 asked respondents if any one of a series of items (B25_a-B25_m) was a reason they were not currently working. Additionally respondents were asked at B26 if there are any other reasons they were not working that had not been mentioned. If the answer was yes, a verbatim response was collected at B27. Prior to coding, verbatim responses at B27 were reviewed to determine if they could be back-coded into B25_a through B25_m, or, if not, whether they could be clustered into additional categories. Table IV.2 provides the response categories added for coding. Responses were then back-coded when possible into one of the existing or newly created categories. Responses that could not be coded were retained as "other." If all responses could be coded, B26 was recoded to "no." If a verbatim response could not be coded into any of the B25 categories, B26 remained coded "yes."

TABLE IV.2

RESPONSE CATEGORIES ADDED TO SECTION B DURING CODING

Item	Question Text	Categories Added
B25	Are you not working because…	n=CAN'T FIND A JOB o=LACK SKILLS
B39	Who {do you/does NAME} discuss your work goals with the most?	10=OTHER NONRELATIVE
B42	Who else {do you/does NAME} discuss {your/his/her} work goals with?	10=OTHER NONRELATIVE
B45	Who else {do you/does NAME} discuss {your/his/her} work goals with?	10=OTHER NONRELATIVE

5. Back-Coding Field-Coded Responses

Items B39, B42, and B45 asked respondents to indicate who they discussed their work goals with. Respondents provided a verbatim response to these questions which interviewers then attempted to code into one of eight response categories during the interview. Responses that were coded as "other" by the interviewer were reviewed by coders and back-coded into existing response options when possible. The additional response option "other nonrelative" was added to each item to capture additional responses (see Table IV.2). Verbatim responses that could not be recoded into one of these 10 categories were left coded as "other."

C. SECTION C—CURRENT EMPLOYMENT

This section collected information about respondents' current job(s). Respondents were asked about the type of work performed, kind of business, hours worked, benefits offered, and wages earned. The section also asked about work-related accommodations received as well as those needed but not received. Other questions gathered information about job satisfaction.

Job specific information (items C2-C13) was collected separately for each current job held. These items are represented in the data file with an _n indicating which job the data are in reference to (for example, C4mth_1 indicating month started first job, C4mth_2 indicating month started second job, and so on). Respondents were asked to report on their main job first (that is the job at which they worked the most hours) and then to subsequently report on other jobs currently held. For the purposes of the constructed variables based on data collected in this section, constructs pertaining to the "main" job are all based on responses provided in the first job slots (_1).

1. Occupation and Industry Coding

Respondents were asked at item C2 to describe the kind of work they did at each of their current jobs (occupation). We used the Bureau of Labor Statistic's 2000 Standard Occupational Classification (SOC) to code verbatim responses to the occupation items.[8] The SOC is a system for classifying all occupations in the economy, including private, public, and military occupations in which work is performed for pay or profit. Occupations are classified on the basis of work performed, skills, education, training, and credentials. The sample member's occupation was assigned one occupation code. The first two digits of the SOC codes classify the occupation to a major group and the third digit to a minor group. For the NBS we assigned three-digit SOC codes to describe the major group the occupation belonged to and the minor groups within that classification (using the 23 major groups and 96 minor). We also assigned three-digit SOC codes to identify the major group comprising the occupation and the minor groups within that classification.

Information about the kind of business where the sample person was employed was collected at C3 (industry). Verbatim responses to the industry items were coded using the 2002 North American Industry Classification System (NAICS).[9] The NAICS is an industry classification system which 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 20 industry sectors. For the NBS, we coded NAICS industries to three digits: the first two numbers specify the industry sector, and the third

---

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

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

number specifies the sub-sector. Both the SOC and the NAICS coding schemes are used in most federal surveys, thus providing uniformity and comparability across data sources. Although both of these classification systems allow coding to a greater level of specificity, based on the research needs of the project, a decision was made with SSA and the analysts to limit coding to the three-digit level. More information on coding responses to the health condition items can be found in "The National Beneficiary Survey: Round 1 Editing, Coding, Imputing, and Weighting Procedures" report (Potter et al. 2008).

The verbatim responses provided at C2 and C3 do not appear on either the restricted or public use version of the file. Rather, the coded responses to C2 for each job listed are found in the constructed variables C_MainCurJobSOC, C_CurJob2SOC, and so on and the coded responses to C4 are found in C_MainCurJobNAICS, C_CurJob2NAICS, and so on.

a.   Uncodeable Occupation and Industry Verbatim Responses

We anticipated that some verbatim responses would lack enough detail to allow coding at the three-digit level. We provided coders with supplemental two-digit codes to allow a general level code to be assigned in these instances (see Table IV.3). In cases where a respondent did not provide a codeable occupation but indicated either in the verbatim response or at C7 (job part of sheltered workshop) that the occupation was a sheltered workshop position, we assigned the code 94. This code was assigned only if the occupation could not be assigned an SOC code. If the position was at a sheltered workshop, but a codeable occupation was provided, the occupation was coded using the SOC classification. When respondents indicated in C7 that their current job was a sheltered workshop position, the industry was coded as 624 (social assistance), which encompasses service for people with disabilities. If the occupation was uncodeable and there was no indication that the position was a sheltered workshop position, the code 99 (uncodeable) was

39

assigned to occupation. In all, less than one percent of the current occupation verbatim responses and one percent of all industry verbatim responses were uncodeable.

TABLE IV.3

TWO-DIGIT SUPPLEMENTAL CODES FOR OCCUPATION AND INDUSTRY CODING

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

2. Hours Worked

At item C8, respondents were asked to provide the number of hours per week usually worked at their current job. A soft edit check was incorporated into the Blaise instrument to prompt interviewers to verify that this response was correct for any response over 60 hours per week. All responses over 60 hours a week (four cases for job one, for example) and under five hours a week (69 cases for job one) were examined during data cleaning. After a review of other job-related information, including occupation and industry verbatim responses, wage rates, self-employment, and sheltered workshop indicators, we concluded that three cases should be recoded. In two cases, it appeared that a zero had been left off in data entry, which was corrected. In the other case, C8 was set to missing (.D) to be imputed later. In general, if the respondent was working in a sheltered employment setting, we determined that low values for hours worked were not unreasonable and should be retained. Similarly if the respondent occupation was consistent with a high number of hours worked per week (for example truck driver), the values

were retained. While some other values were suspect, in general, our approach was to only recode cases that appeared to be obvious data entry or respondent errors.

3. Weeks Per Year

At item C9, respondents were asked how many weeks per year they usually worked at their current job. Responses of fewer than 20 weeks were examined during data cleaning (70 cases). Other job-related information was reviewed in an attempt to determine whether the values were reasonable. In two cases, cases were recoded as "not working" for B24 based on notes indicating that the respondent was, in fact, not currently working. In general if the occupation verbatim and other job-related information was consistent with the possibility of minimal weeks worked per year, the original values were retained. In some cases, it appeared that respondents had interpreted the question to be asking how many weeks they had worked if they had just started their job (despite the inclusion of the probe "If you have worked less than a year, please answer for the number of weeks you expect to work."). Since it was not possible based on other information to determine whether these values were errors, they were retained on the data file.

4. Pay

Respondents were asked to report their pretax earnings for each current job at C11 (if reported as an hourly wage) or C12amt (if reported in another unit, such as daily, weekly, monthly, or annually) and their take home pay at C13amt. Three constructed variables were created, one designed to combine pretax responses into an hourly wage (C_MainCurJobHrPay, C_CurJob2HrPay, and so on) and one into a monthly wage (C_MainCurJobMnthPay, C_CurJob2MnthPay, and so on) regardless of where the initial reporting occurred. A construct was also created for monthly take home pay (C_MainCurJobPayTH, C_CurJob2MnthPayTH, and so on). In addition, a total monthly pay variable was created to sum across all jobs

(C_TotCurMnthPay). Because the earning constructs were subject to imputation, there was concern that outliers might become imputation donors and exacerbate the outlier problem. This concern prompted a detailed review of high and low values for both the source variables and constructs. Cases with very high and very low values were excluded from the donor pool for imputation.

A soft edit check was included in the Blaise instrument to prompt interviewers to verify any response over $25 an hour at C11. This check could be suppressed, however, which resulted in six cases reporting hourly rates over $25 an hour. Since other job-related information, including the verbatim occupation response, indicated that these could be valid entries, all were retained on the file. Hourly wage values of $3 and below were also examined. In these cases, since the respondent was working in a sheltered employment setting or the verbatim job description indicated that the low values for hourly wages was not unreasonable, these values were retained.

Soft edit checks were also built into the instrument to flag high entries for each of the various reporting units at C12amt and C13amt. Values that were suppressed or that were at the high and low end of the range were examined. In most cases, the verbatim occupation and industry descriptions indicated that the values could be valid, thus they were retained on the file. Generally, if the respondent was working in a sheltered employment setting or the verbatim job description indicated that the low values for wages were not unreasonable, these values were retained. In four cases, where it seemed clear that an interviewer or respondent error had been made, C12amt was set to missing (.D), and in two cases C13amt was set to missing to be imputed later.

In addition to examining high and low values, take-home and pretax values were compared. While some cases (78) had a difference of 30 percent or more and were initially flagged for verification, only those with the most extreme differentials whose other job-related information

did not support the difference were recoded or set to missing (three cases). In four cases where the take-home pay reported was higher than the pretax pay, C12amt was set to missing. In four additional cases, it appeared that the interviewer had made an error in coding the unit at C13hop or C12hop, which was corrected. In future rounds, we recommend including a soft edit check within the instrument to prompt interviewers to validate responses with a large discrepancy between pretax and take-home pay.

While many questionable values remain on the file, two flag variables were created and are included on the file to identify cases reporting total monthly pay over $10,000 and cases reporting pay less than $20 monthly or $1.50 hourly. Users of the data file may choose to eliminate these cases from analyses.

5. Amount Paid for Personal Assistance Service

Respondents who indicated that they received equipment or personal assistance to help work were asked in items C26amt and C31amt respectively, how much they paid for such services. Respondents were given the opportunity to respond in a variety of ways including how much was paid per week, per month, per year, or whether the service was paid in a one-time payment. The constructs C_CurMnthEquipExp (Monthly Equipment Expenses) and C_CurMnthPASExp (Monthly PAS expenses) were created to consolidate responses to a monthly amount. However, it was later determined that one-time payments should not have been an option for reporting the amount paid for personal assistance services. While monthly amounts were calculated for the five cases reporting one-time payments (by amortizing over five years), this response option should be dropped from item C31hop in future rounds.

43

6. Back-Coding Responses to "Other/Specify" Items

Items C33a-e asked whether a series of accommodations were made by the sample member's employer. If the respondent indicated that other accommodations were made (C33_f=1), a verbatim response was collected. These responses were reviewed and back-coded into C33_a-e when possible.

Respondents were also asked whether changes were needed but not made to the sample member's workplace (C34). If yes, a verbatim response was collected to get specific information about what changes were needed at C35. The verbatim responses were reviewed prior to coding, and five categories were created to summarize responses (see Table IV.4). Responses that could not be coded into one of these five categories, were retained as "other."

TABLE IV.4

RESPONSE CATEGORIES ADDED TO SECTION C ITEMS DURING CODING

Item	Question Text	Categories Added
C23	What kind of special equipment {do you/does NAME} use?	7=HEARING AIDS 8=GLASSES
C35	Are there any changes in {your/NAME's} {main/current} job or workplace related to {your/his/her} mental or physical condition that {you need/(he/she needs}, but that have <u>not</u> been made? (IF YES) What are those changes?	a=NEED SPECIAL EQUIPMENT b=NEED CHANGES IN SCHEDULE c=NEED CHANGES TO THE TASKS d=NEED CHANGES TO ENVIRONMENT e=NEED CO-WORKERS TO ASSIST

7. Back-Coding Field-Coded Responses

Item C23 (what kind of special equipment was used at work), C24 (who paid for equipment used at work), and C28 (what kind of personal assistance services is used at work) were all open-ended items that interviewers attempted to code into one several predefined response categories during the interview. Responses that were coded as "other" by the interviewer were reviewed by coders and back-coded into existing response options when possible. For item C23, additional

response options were added to capture additional "other" responses (see Table IV.4). Verbatim responses that could not be recoded into one of these 10 categories were left coded as "other."

D. SECTION D—JOB/OTHER JOBS DURING 2003

This section collected information about employment during the 2003 calendar year, including type(s) of employer(s), hours worked, wages earned, and reasons for leaving employment, if applicable. Other questions asked if respondents worked or earned less than they could have (and if so, the reasons why), and collected information about experiences related to Social Security benefit adjustments due to work.

As in section C, job specific information (items D2-D23) was collected for each job held in 2003. Data for each job are represented on the data file with an _n indicating which job the data are in reference to (for example, D6mth_1 indicating month started first job, D6mth_2 indicating month started second job, and so on). 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 2003 that had previously been mentioned in section C as current employment. Rather, employment data from section C was copied to section D items during data processing for all current jobs also held during the 2003 time period. See Table IV.5 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 (D8mnth, D8yr, D23, D23_oth) were coded as .L (indicating logical skip).

TABLE IV.5

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

1. Including Current Jobs Held in 2003 in Section D

Jobs mentioned in section C were defined as held in 2003 if C4yr (year started current job) was earlier than or equal to 2003. 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 data file to indicate which jobs from section C (by job number) were copied into specific section D job slots.

2. Determining the Main Job Held in 2003

In addition to copying job data from section C to the section D items, it was necessary to determine which job held in 2003 was the main job. Prior to including the jobs from section C, the main jobs held in 2003 were stored as job 1. Since it was possible that a job reported in section C was the respondent's main job in 2003, hours worked in 2003 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 2003 job[10]. The variable Main_Job_grid_num identifies the job number of the main job held in 2003 after this analysis. 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 2003 (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 2003.

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 "main" job (C_MainJob2003SOC, C_MainJob2003NAICS, C_MainJobHrPay2003, C_MainJobMnthPay2003, C_MainJobMnthPayTH2003, and C_MnthsMain2003Job) 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.

---

[10] 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 2003 section C jobs, job 1 in section D was counted as the main job in 2003. If there were no jobs listed in section D, and hours per year was missing for all 2003 jobs in section C, the first job listed in C that was a 2003 job was counted as the main job in 2003. 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 2003 job. If there was no 2003 job from section C, or hours per year was missing for all section C 2003 jobs, job 1 in section D was counted as the main 2003 Job. If hours per year was missing for all 2003 section C jobs and job 1 in section D, job 1 in section D was counted as the main job in 2003.

During data processing, we found 38 cases in which the respondent reported in B30 that he or she did not work in 2003 (B30=0) but whose reported current job start dates indicated that a job was held in 2003. These cases were recoded to B30=1 (indicating did work in 2003). Note that D3 ("Other than the current jobs you just told me about, how many other jobs did you hold for at least one month in 2003?") was not recoded to reflect the number of jobs held in 2003 after including jobs from section C.

3. Occupation and Industry Coding

Respondents were asked at item D4 to describe the kind of work they did on each of the jobs they held in 2003 (occupation) and at item D5, to describe the kind of business (industry). As for the equivalent items in section C, the verbatim responses to these items were coded using the SOC and NAICS classification systems described above. The verbatim responses to D4 and D5 are not provided on the restricted or public use version of the data file. Rather the coded responses to D4 are found in the construct C_MainJob2003SOC, C_Job12003SOC, and so on. The coded responses to D5 are found in C_MainJob2003NAICS, C_Job12003NAICS, and so on.

a. Uncodeable Occupation and Industry Verbatim

Coders used the same supplemental two-digit codes described above to assign general level codes when full SOC and NAICS could not be assigned. In all, less than one percent of the 2003 occupation verbatim responses and one percent of all industry verbatim responses were uncodeable for any given item.

4. Dates Worked at 2003 Job

Items D6mth, D6yr, D8mth, and D8yr collected start and stop dates for each job held in 2003. Soft edit checks were built into the Blaise instrument to verify that stop dates were later than start dates and to verify that each job was held for at least one month in 2003. If the

interviewer verified that the job ended before 2003 or that the job was held for less than one month in 2003, items collecting job specific information at D14-D21hop were skipped. Occupation and industry data as well as start and stop dates for these jobs are maintained on the data file because respondents were asked other items in section D (why they stopped working at the job at D23 and general items about working in 2003 at D25-D30).

5. Hours Worked

At item D16, respondents were asked to provide the number of hours per week usually worked at their 2003 job. As in section C, a soft edit check was incorporated into the Blaise instrument to prompt interviewers to verify that this response was correct for any response entered over 60 hour per week. Responses over 60 hours a week (10 cases for job one for example) and under 5 hours a week (22 cases on job one) were examined during data cleaning. After reviewing other job related information, data from all cases were retained. In general, if the respondent was working in a sheltered employment setting, we determined that low values for hours worked were not unreasonable and should be retained. Similarly if the respondent's occupation was consistent with a high number of hours worked per week the values were retained.

6. Weeks per Year

At item D17 respondents were asked how many weeks per year they usually worked at their 2003 job. Responses less than 20 weeks were examined during data cleaning (210 cases for job one). As for hours, other job-related information was reviewed in an attempt to determine whether the values were reasonable. In general if the occupation verbatim response and other job-related information was consistent with the possibility of few weeks worked per year, the original values were retained. As in section C, in some cases, particularly for those who had

recently started a job, it appeared that respondents had interpreted the question to ask how many weeks they had worked despite the inclusion of the probe "If you have worked less than a year, please answer for the number of weeks you expect to work." Since it was not possible based on other information to determine whether these values were in fact errors, they were retained on the file.

7. Pay

Respondents were asked to report their pretax earnings for each current job at D19 (if reported as an hourly wage) or D20amt (if reported in another unit, such as daily, weekly, monthly, or annually) and their take-home pay at D21amt. Three constructed variables were created, one designed to combine pretax responses into an hourly wage (C_MainJobHrPay2003, C_Job1HrPay2003, and so on) and one into a monthly wage (C_MainJobMnthPay2003, C_Job1MnthPay2003, and so on) regardless of where the initial reporting occurred. A construct was also created for monthly take-home pay (C_MainJobMnth PayTH2003, C_Job1Mnth PayTH2003, and so on). In addition, a total monthly pay variable was created to sum across all jobs (C_Tot2003Pay). Source variables and later constructed variables were examined for extremely high and low values.

A soft edit check was incorporated into the Blaise instrument to prompt interviewers to verify any response over $25 an hour at D19. Responses over $25 an hour (three cases for job one) were closely examined. Since other job-related information, including the verbatim occupation response, indicated that these could be valid entries, all were retained on the file. Hourly wage values of $3 and below were also examined. In these cases, because respondents were working in a sheltered employment setting or the verbatim job description indicated that the low values for hourly wages were not unreasonable, these values were retained.

Soft edit checks were also built into the instrument to flag high entries for each of the various reporting units at D20amt and D21amt. As for hourly wages, values that were suppressed or that were at the high and low end of the range were examined. Cases for each reporting unit were examined by looking at other job-related information. In most cases, the verbatim occupation and industry descriptions indicated that the values could be valid, thus they were retained on the file. Generally, if the respondent was working in a sheltered employment setting or the verbatim job description indicated that the low values for wages were not unreasonable, these values were retained. In one case, where it seemed clear that an interviewer or respondent error had been made, D20amt was set to missing (.D). Recoding of data occurred only when there was an obvious data entry error or when the respondent's job characteristics were not consistent with reported earnings or pay.

In addition to examining high and low values, take-home and pretax values were compared. While many cases (70 for job one) had a difference of 30 percent, only those with the most extreme differentials and whose other job-related information did not support the difference were recoded or set to missing (two cases). In four cases where the take-home pay reported was higher than the pretax pay, D20amt or D21amt was set to missing (.D). In future rounds, we recommend including an edit check within the instrument to prompt interviewers to validate responses with a large discrepancy between pretax and take-home pay.

8. Back-Coding Responses to "Other/Specify" Items

D25_a-D25_f asked if a series of issues were reasons the sample person had worked fewer hours than they could have. Item D26 asked if any of D26a-D26h were reasons the sample member did not work or earn more. Responses coded as "other" were reviewed during data processing. For both items, an additional category was added during coding to allow further categorization of responses (see Table IV.6).

TABLE IV.6

RESPONSE CATEGORIES ADDED TO SECTION D DURING CODING

Item	Question Text	Categories Added
D23	Why did {you/NAME} stop working at this job?	19=MOVED TO ANOTHER AREA 20=FOUND ANOTHER JOB 21=LOSS OF BENEFITS 22=WORK SCHEDULE
D25a	Did you work fewer hours or earn less money than you could have because you…	g=HAD MEDICAL PROBLEMS
D26	In 2003, do you think {you/NAME} could have worked or earned more if {you/he/she} had:	i=BETTER HEALTH/TREATMENT

9. Back-Coding Field-Coded Responses

Item D23 asked why the sample person quit working at the job held in 2003. Interviewers attempted to code the verbatim responses into a series of predetermined categories if possible. Cases coded as "other reason" by interviewers were reviewed to determine if they could be back-coded into an existing category. In addition, after reviewing the verbatim responses, four additional categories were created to facilitate coding (see Table IV.6). Responses that could not be coded into one of these five categories were retained as "other."

E. SECTION E—AWARENESS OF SSA WORK INCENTIVE PROGRAMS AND TICKET TO WORK

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

1. Missing Response on Awareness of Work Incentive Items

There were two problems in Section E that resulted in eligible respondents being skipped for certain questions they should have received. All SSI beneficiaries were to be asked questions

E3-E10 (awareness and use of PASS, Earned Income Exclusion, PESS, and Continued Medicaid Eligibility); SSI beneficiaries who were 25 or younger at time of sampling (September 2003) and who received SSI benefits before age 22 were to be asked questions E12-E13 (awareness and use of Student Earned-Income Exclusion); and all DI beneficiaries were to be asked questions E15-E18 (awareness and use of Trial Work Period and Extended Period of Eligibility for Medicare). Due to a specification error, beneficiaries who received both SSI and SSDI incorrectly skipped these questions. A second problem occurred when a respondent was unable to complete the interview in one session. When the interview continued in a second session, the information about the sample member's SSI and DI status, age at interview, and sample member type (Ticket Participant sample or Representative Beneficiary Sample) was not saved in the system. This also resulted in Ticket to Work participants who said that they had not heard of the Ticket to Work program in items E21 and E22 skipping E24 ("Are you aware that according to Social Security, you are participating in the Ticket to Work program and your Ticket is assigned to {EN} as of {date}?"). Once these problems were discovered, they were quickly corrected. Table IV.7 provides a summary of the number of respondents who skipped items for which they were eligible. In the data file, those that incorrectly skipped the item are coded as .M (missing due to error). In addition, the variables E2Skip (Skipped E3 to E10), E12skip (Skipped E12 to E13), E15skip (Skipped E15 to E18), and E24skip (Skipped E24) were created to allow analysts to identify which respondents inappropriately skipped these items. A value of 1 on these items indicates that respondent was eligible to receive the item, but it was not administered.

TABLE IV.7

QUESTIONS E3, E12 AND E15 SKIP PATTERNS

	Questions			
	E3	E12	E15	E24
Total	7,603	7,603	7,603	7,603
Asked	1,796	440	1,898	65
Incorrectly Skipped	2,762	579	2,728	45
Logically skipped	3,045	6,584	2,977	7,493

Note: E3 asks the respondent if the beneficiary (who might be the respondent) has ever heard of a <u>Plan for Achieving Self-Support or a PASS Plan.</u> This is a Social Security incentive that lets beneficiaries set aside money to be used to help them reach a work goal. The money set aside does not affect their benefits. E12 asks the respondent if the beneficiary (who might be the respondent) has ever heard of the <u>student earned-income exclusion.</u> This is a Social Security incentive where if a beneficiary is in school, up to $1,340 of earnings per month are not counted when Social Security figures the benefit. E15 asks the respondent if the beneficiary (who might be the respondent) has ever heard of a <u>Trial Work Period</u>. This is a Social Security incentive that lets beneficiaries earn above $800 per month for nine months without losing their benefits. E24 asks if the respondent is aware that the beneficiary is participating in the Ticket to Work Program if the respondent has indicated being unaware of the program.

Because these items are particularly important to SSA, a special set of weights was calculated to account for the missing data on items E3-E10, E12-E13, and E15-E18: WGT1_COMBFINLE3 (final combined Weight for E3 skip), WGT1_COMBFINLE12 (final combined Weight for E12 skip), and WGT1_COMBFINLE15 (final combined Weight for E15 skip). These weights should be used in lieu of the other sample weights on the file to provide national estimates for these particular items. Because E24 was only used to prompt Ticket to Work Participants to recall the name of the program, revised weights were not calculated for this item. Potter et al. (2008) provides more information regarding the calculation of the section E weights.

Nonresponse adjustment procedures are designed to reduce the potential for nonresponse bias, but no nonresponse adjustment procedure can fully remove the potential for nonresponse bias. Because of the lower response rates to Section E, the potential for nonresponse bias is greater for Section E estimates than for estimates based on the responses to other section of the

NBS questionnaire. Estimates based on the section E weights will also have a larger sampling variance than other items because there are fewer sample members with responses.

In subsequent rounds of survey administration, longitudinal respondents who inappropriately skipped E2-E10, E12-E13, and E15-E18 and are re-interviewed will be asked these items.

2. Dates Receiving Services from ENs

There were several problems with the collection of start and stop dates for Employment Networks reported in section E. Respondents were first asked if they were signed up with any Employment Networks (EN) in 2003. If the respondent was no longer signed up with the EN, the month and year the sample member stopped receiving services was collected in E42mth and E42yr. Twenty-four respondents reporting leaving the first EN mentioned before 2003; two gave dates before 2003 for the second EN, and one gave a date before 2003 for the third EN mentioned. In these cases, E42yr was set to missing (.D).

Respondents who had not reported being currently signed up with an EN that they were signed up with in 2003 (E41) were asked if they were currently signed up with any EN at E45. The month and year the sample member started receiving services from this EN was collected at E47mth and E47yr. Although some respondents reported first receiving services in 2003 or earlier (32 cases), E37 ("Were you signed up with any Employment Network at any time in 2003?") was not recoded to "yes" for these cases since recoding this item would have affected the skip logic in section H. Receiving questions in section H related to Ticket use in 2003 were based on E37 and E40_yr only. Additionally, respondents who had not yet reported ever using a Ticket with any other ENs were asked what month and year they first began receiving services at E50mth and E50yr. While nine cases reported receiving first signing up in 2003, E37 was not recoded during data processing. These inconsistencies remain on the data file.

Finally, in three cases, respondents provided start dates with their current EN that overlapped with stop dates reported with other ENs. Since Tickets can only be assigned to one EN at a time, these are improbable values. While the dates provided in E42mth, E42yr, and E47mth and E47yr were retained, for the purposes of the construct C_TotMnthsTTW (Total Months Enrolled in TTW in 2003), these cases were top-coded at 12.

3. Multiple Current ENs

There were also 17 cases in which the respondent reported being currently signed up with more than one EN at E41. Since this was a relatively rare problem and it was unclear which EN should be considered current, these inconsistencies were not recoded and remain on the file.

4. Back-Coding Field-Coded Responses

Item E28, E30, and E32 asks who sent, called, or talked to the sample person about the Ticket to Work program. Interviewers attempted to code the respondent's verbatim response into one of 10 response categories. Responses that interviewers coded as "other" were examined and back-coded when possible. Responses that could not be back-coded were retained as "other."

5. Coding Open-Ended Responses

Item E43 was an open-ended question that asked respondents why they were no longer receiving services from their 2003 EN. Before coding, the verbatim responses were reviewed and five categories were created to cluster responses (see Table IV.8). Responses that could not be coded into one of these categories were retained as "other."

TABLE IV.8

RESPONSE CATEGORIES ADDED TO SECTION E DURING CODING

Item	Question Text	Response Categories Added
E43	Why {are you/is NAME} no longer receiving services from {EN IN 2003 FROM E39}?	1=NEVER RECEIVED ANY INFO 2=FOUND A JOB 3=CANNOT WORK FOR HEALTH REASONS 4=OTHER REASON RELATED TO PERSONAL CIRCUMSTANCE 5=OTHER REASON RELATED TO EN 6=OTHER

F. SECTION F—TICKET NONPARTICIPANTS IN 2003

This section collected information about reasons for nonparticipation in the TTW program. It asked whether the respondent had attempted to learn about employment opportunities (including TTW), problems he or she may have had with Employment Networks or other employment agencies, and how those problems were handled or resolved.

1. Back-Coding of Other/Specify Responses

Question F2 asked if the sample member contacted any of a series of agencies or individuals to get information about TTW. Two "specify" response options (an Employment Network and Other Agency or Organization) prompted a verbatim response. During data processing, the verbatim responses were reviewed and back-coded into F2_a-F2_g when possible.

2. Back-Coding Field Coded Responses

Items F6, F8, and F10 ask who sent information, called, or talked to the sample person about the Ticket to Work program. Item F29 asks for reasons the sample person did not contact the State VR after receiving information. Responses that interviewers coded as "other" were examined and back-coded when possible. Responses that could not be back-coded were retained as "other." An additional response category was added to item F29 (to assist with back-coding of other/specify responses post processing (see Table IV.9).

TABLE IV.9

RESPONSE CATEGORIES ADDED TO SECTION F DURING CODING

Item	Question Text	Response Categories Added
F14	Why didn't {you/NAME or his/her representative} try to use {your/NAME's} Ticket with the State VR agency in 2003?	1=AGENCY DIDN'T HELP 2=DID NOT KNOW COULD 3=WAS NOT HEALTHY ENOUGH 4=OTHER
F29	After receiving information about the Employment Networks in {your/NAME's} area including the State VR agency or {STATE NAME FOR VR}, why didn't {you/NAME or his/her representative} contact any of them?	15=GOT A JOB OR IN SCHOOL
F31	What are the main reasons {you did/NAME did} not try to participate in the Ticket to Work program in 2003?	1=HEALTH REASONS 2=HAD A JOB/IN SCHOOL 3=DID NOT KNOW ABOUT PROGRAM 4=DID NOT WANT TO/DID NOT TRY 5=OTHER 6=CANNOT WORK, REASON UNSPECIFIED

3. Coding Open-Ended Responses

F14 (reasons did not try to use the Ticket with a State Vocational Rehabilitation Agency in 2003), F23 (reasons did not try to use the Ticket with the EN contacted), and F31 (reasons did not try to participate in TTW in 2003) were all open-ended item to which respondents provided a verbatim response. Based on a review of the responses, categories were developed based on common responses (see Table IV.9). Coders then attempted to code the verbatim response into an established category. If the response did not fit into one of these categories, it remained as "other." Because there were too few responses to F23 (24 cases), the responses could not be clustered and assigned numeric codes.

G. SECTION G—EMPLOYMENT-RELATED SERVICES & SUPPORT USED IN 2003

This section collected information from respondents about their use of employment-related services in 2003, 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 asked about sources of information about services and the nature of any services that were needed but not received.

1. Missing Provider Names

To aid in the recall of employment-related services received in 2003, 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. In several cases, respondents did not know the name of the provider so the provider type followups were not asked.[11] Additionally, in some cases, interviewers left the first provider slot empty, either in error or because the respondent did not know the name of the provider. This also caused the provider type follow-up items to be skipped. In these cases, provider name was set to Don't Know (.D). This affected 19 cases at G2, 55 cases at G11, 25 cases at G16, and 124 cases at G20. In total, nine percent of provider names were missing at item G2 and G11, seven percent were missing at G16, and six percent were missing at G20. We recommend in subsequent rounds disabling the option to code a provider name as "Don't Know" and emphasizing the importance of collecting some information under provider name, even if it is a description of the place rather than the name.

To minimize respondent burden by avoiding the need to ask provider type again if a provider type was listed under two or more services, interviewers could indicate that a provider

---

[11] In the data file, if an interviewer selected the "don't know" response, it appears as 99999999999999999999998 and 99999999999999999999999 if "refused".

59

had already been mentioned, thus skipping the provider type follow-up questions. In some cases however, interviewers indicated that a provider had already been mentioned, when in fact it had not been. This resulted in missing data on the provider type questions. Cases where providers were inappropriately deleted were identified by careful examination of Section G data. For example, we examined cases in which G1=0 (no employment services received), G10=1 (received job training), but where G13_1=.L. In these cases, the interviewer had indicated at G12 that the first provider given at G11 had already been mentioned (causing the provider type followups to be skipped), which was not possible. In cases such as this, the provider type items (G13 and G14) were set to .M indicating an error caused the item to be skipped. Similar strategies were employed to examine providers marked as already mentioned at G17 and G20. In all, there were 23 cases in which the provider type items G13 and G14 were set to missing (.M) for a provider listed at G11, 97 cases in which provider type item G18 was set to missing (.M) for a provider listed at G16, and 104 cases in which provider type item G22 was set to missing (.M) for a provider listed at G20. We recommend that in future rounds the screens requiring interviewers to indicate if a provider has already been mentioned be simplified and increased time be spent during training on this procedure so that there is less likelihood of providers being mistakenly marked as duplicates.

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 2003 were asked about each provider from whom services were received in 2003. Before asking when services were received, the list of providers listed at items G2, G11, G16, and G20 was compiled for the interviewer, and the interviewer was asked to verify if any of the providers on the list were duplicates. The interviewer was also asked to verify with the respondent if any of the

providers on the list were the same. Providers marked as duplicates were removed from the list and the provider specific follow-up information was not obtained. While this process worked relatively well, there were some cases in which providers were marked as duplicates that did not appear to be duplicates based on an examination of provider name and type. These cases were coded as .M (indicating missing due to error) on item G33 since the followups regarding when services were received were not asked of these providers. Across providers, there were 456 instances in which this data was missing. For any single provider, the first provider listed under mental health services at G19 had the highest proportion of missing data with 6.4 percent (152 cases) cases coded as .M. The restricted access file includes a flag variable for each provider indicating whether the provider was marked for removal from the list (e.g. G_Del_1-G_DEL_34). Cases in which the provider name was coded as missing in G2, G11, G16, and G20, were not asked G33 or subsequent followups regarding 2003 services (coded as .L=logical skip).

2. Coding Duration of Service Session

At round 1, respondents were allowed to report the length of their service session with each provider in terms of minutes, hours, days, or to indicate that "it varied." Because responses to "it varied" could not be combined with the other categories to create a duration in hours constructed variable (C_DurProvVisit), we recommend dropping this option in future rounds and prompting respondents to give a "usual" estimate if possible.

3. Back-Coding Responses to "Other/Specify" Items

Each of the provider type questions in section G (G7 and G9, G13 and G14, G18, and G22) included an "other" option which prompted a verbatim response. During data processing, the verbatim responses were reviewed to determine whether they could be clustered into additional

categories. Table III.10 provides the response categories added during coding. Responses were then back-coded when possible into one of the existing and newly created categories. Responses that could not be coded were retained as "other." Cases that were back-coded as "state agency" at item G7 were also recoded at G9 to indicate the type of state agency. Cases back-coded as "state agency" at G13 were also recoded on item G14.

Additionally, "other" responses at G36_a-G36_m were reviewed. No additional categories were added during coding, but responses were reviewed and back-coded into existing response options when possible.

4. Back-Coding Field-Coded Responses

Items G28 (type of degree working toward), G45 (who paid for services), G53 (reasons used services), G55 (who pressured to use services), and G56 (how pressured to use services) were all open-ended items that required interviewers to attempt to code the respondent's verbatim response into a predetermined category. Responses that were coded as "other" by the interviewer were reviewed by coders and back-coded into existing response options when possible. In some cases, additional categories were added during coding to cluster "other" responses that did not fit into a predetermined category (see Table IV.10 for categories added).

5. Coding Open-Ended Items

Item G61 (reasons unable to get services needed) was an open-ended question with no response options specified. Based on a review of the responses, seven categories were developed based on common responses (see Table IV.10). Coders then attempted to code the verbatim response into an established category. If the response did not fit into one of these categories, it remained as "other."

TABLE IV.10

RESPONSE CATEGORIES ADDED TO SECTION G DURING CODING

Item	Question Text	Response Categories Added
G7	Thinking about {PROVIDER FROM G2}, was this place:	4=SCHOOL
G9	Thinking about {PROVIDER FROM G2], was this place:	6=A WORKFORCE CENTER/ EMPLOYMENT OFFICE
G13	Thinking about {NEW PROVIDER FROM G11}, was this place:	4=A SCHOOL OR COLLEGE
G18	Thinking about {NEW PROVIDER FROM G16}, was this place:	5=A SCHOOL 6=A NURSING HOME/GROUP HOME 7=A GOVERNMENT AGENCY 8=IN HOME CARE 9=A MEDICAL EQUIPMENT STORE 10=A REHABILITATION CENTER 11=PHYSICAL THERAPY CENTER
G22	Thinking about {NEW PROVIDER FROM G20}, was this place:	6=RESIDENTIAL TREATMENT 7=REHAB CENTER
G45	In 2003, who paid for the services {you/NAME} received from {PROVIDER FROM G32 DE-DUPLICATED LIST IF USED IN 2003}?	14=SCHOOL/FINANCIAL AID/GRANT 15=STATE AGENCY/COUNTY/GOVERNMENT
G53	Thinking only about services {you/NAME} used in 2003, what are the main reasons you decided to use these services?	9=TO BE MORE INDEPENDENT
G55	Who pressured {you/NAME} to use these services?	13=HEALTH PROVIDER 14=COURT/POLICE
G56	How did {your/NAME's} {FILL PERSON(S) FROM G55} pressure {you/him/her} to use these services?	6=THREATENED HOSPITALIZATION/JAIL
G61	Why {were you/was NAME} unable to get these services?	1=NOT ELIGIBLE/REQUEST REFUSED 2=LACK INFORMATION 3=COULD NOT AFFORD 4=DID NOT TRY 5=TOO DIFFICULT/TOO CONFUSING 6=PROBLEMS WITH THE SERVICE 7=OTHER

H. SECTION H—TICKET PARTICIPANTS IN 2003

This section asked 2003 TTW participants about their experiences with the program, including information related to how they decided to participate in the Ticket program, the kinds of information they used to pick their current service providers, development of their individual work plan (IWP), and any problems experienced with services provided by an Employment Network. The section also included a series of questions about how problems with Employment Networks were resolved and overall satisfaction with the TTW program.

1. Back-Coding Responses to "Other/Specify" Items

Question H7 asked respondents if they contacted any of a series of agencies or individuals to get information about the TTW program. Two "other" response options (Other Agency and Anyone Else) prompted a verbatim response. During data processing, the verbatim responses were reviewed and back-coded into H7_a-H7_h when possible.

2. Back-Coding Field Coded Responses

Items H14 (who sent information about ENs), H16 (who called to talk about ENs), H18 (who talked to about ENs), H25 (reasons State VR did not accept ticket), H31 (reasons EN did not accept ticket), H35 (reasons chose EN), H50 (what did to try to solve problems), H52 (who gave information about getting help with problem), and H60 (why did not try to solve problem) were all open-ended items that required interviewers to attempt to code the respondent's verbatim response into a predetermined category. Responses that were coded as "other" by the interviewer were reviewed by coders and back-coded into existing response options when possible. In some cases, additional categories were added during coding to cluster "other" responses that did not fit into a predetermined category (see Table IV.11 for categories added).

TABLE IV.11

RESPONSE CATEGORIES ADDED TO SECTION H AS A RESULT OF CODING

Item	Question Text	Response Categories Added
H3	Why did {you/NAME} decide to participate in the Ticket to Work program?	1=WANTED TO GET A JOB/ MORE MONEY 2=WANTED TO FEEL MORE INDEPENDENT 3=OTHER
H23	Why didn't {you/NAME or his/her representative} try to use {your/NAME's} Ticket with the State VR agency in 2003?	1=SIGNED UP WITH OTHER AGENCY 2=ALREADY RECEIVING SERVICES 3=OTHER
H29	Why didn't {you/NAME or (his/her) representative} try to use {your/NAME's} Ticket with {any of} the other Employment Network(s) {you/NAME or (his/her) representative} contacted in 2003?	1=LOCATION 2=OTHER
H31	Why didn't {any of} the other Employment Network(s) {you/NAME} tried to use {your/his/her} Ticket with accept {your/NAME's} Ticket in 2003?	7=TROUBLE CONTACTING EN
H33	What information did {you/NAME} need but didn't get?	1=HOW/WHERE TO USE THE TICKET 2=SERVICES PROVIDED 3=OTHER
H35	Why did {you/NAME or (his/her) representative} choose {{LONGEST} EMPLOYMENT NETWORK IN 2003}?	8=KNEW ABOUT THEM OR REFERRED TO 9=FINANCIAL COMPENSATION
H38	What problems did {you/NAME} have during 2003 (with the services you received from EN)?	1=PROBLEMS MAKING CONTACT 2=PROBLEMS NOT RECEIVING SERVICES 3=PROBLEMS WITH COUNSELOR 4=OTHER
H48	What was the problem about?	1=PROBLEMS MAKING CONTACT 2=PROBLEMS RECEIVING SERVICES 3=OTHER

3. Coding Open-Ended Items

Items H3 (reasons decided to participate in TTW), H23 (reasons did not try to use ticket with State VR), H29 (reasons did not try to use ticket with other ENs), H33 (what information needed but didn't get), H38 (what problems had with EN), and H48 (what problem with EN was about) were open-ended questions with no response options specified. Based on a review of the responses, categories were developed for each item based on common responses (see Table

IV.11). Coders then attempted to code the verbatim response into an established category. If the response did not fit into one of these categories, it remained as "other."

I. SECTION I—HEALTH AND FUNCTIONAL STATUS

This section collects information about the respondent's health status and everyday functioning, including the need for special equipment or assistive devices. Information regarding general health status, difficulties with Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs), a variety of functional limitations, substance abuse/dependence, and treatment for mental health conditions was also collected.

1. Difficulty Seeing

Item I17 asked respondents whether they have any difficulty seeing words and letters in newsprint even when wearing glasses or contacts. If the respondent answered, "yes," "don't know," or "refused," he or she was asked I18, "Are you able to see the words and letters in ordinary newsprint at all?" Respondents who reported difficulty seeing were also asked whether they used any special equipment to help them see (I19), and, if so, what that equipment was (I20). Question I20 was an open-ended item that required interviewers to attempt to code the respondent's verbatim response into a predetermined category. In several cases, respondents reported that the special equipment used was glasses or bifocals (coded as "other" by interviewers). Since the question was meant to identify seeing problems when wearing glasses if usually worn, such responses were considered invalid. If glasses were the only special equipment mentioned, I19 was recoded as "no" and I20 as .L (logical skip).

Given the number of respondents reporting glasses in I20 and feedback from interviewers suggesting respondents did not understand I17, we recommend revising this series in subsequent

rounds so that respondents are first asked whether they wear glasses or contacts and then asked if they have difficulty seeing.

2. Back-Coding Responses to "Other/Specify" Items

Question I20 (equipment used for seeing), I24 (equipment used for hearing), I28 (equipment used for speaking), and I32 (equipment used for walking) were all open-ended items that required interviewers to attempt to code the respondent's verbatim response into a predetermined category. Responses that were coded as "other" by the interviewer were reviewed by coders and back-coded into existing response options when possible. In some cases, additional categories were added during coding to cluster "other" responses that did not fit into a predetermined category (see Table IV.12 for categories added).

TABLE IV.12

RESPONSE CATEGORIES ADDED TO SECTION I AS A RESULT OF CODING

Item	Question Text	Response Categories Added
I20	What devices, equipment, or other types of assistance {do you/does NAME} use? Anything else?	8=MAGNIFYING GLASS
I32	What devices, equipment, or other types of assistance {do you/does NAME} use? Anything else?	9=SPECIAL SHOES OR INSERTS

J. SECTION J—HEALTH INSURANCE

Questions in this section collected information about the sources of health insurance coverage the beneficiary had, both at the time of the interview and during calendar year 2003.

Back-Coding Responses to "Other/Specify" Items. Item J6 (type of private insurance), J9 (type of health coverage), J11 (type of health coverage in 2003) were all open-ended items that required interviewers to attempt to code the respondent's verbatim response into a predetermined category. Responses that were coded as "other" by the interviewer were reviewed by coders and

back-coded into existing response options when possible. For J11, one additional category was added during coding to cluster "other" responses that did not fit into a predetermined category (see Table IV.13 for the category added).

TABLE IV.13

RESPONSE CATEGORIES ADDED TO SECTION J AS A RESULT OF CODING

Item	Question Text	Response Categories Added
J11	Now, I'd like you to think back to 2003. In 2003, what kinds of health coverage did {you/NAME} have?	11=PRIVATE INSURANCE, NOT SPECIFIED WHO THROUGH

K. SECTION K—INCOME AND OTHER ASSISTANCE

Questions in this section asked about sources of income, including income received from earnings, Social Security, Workers' Compensation, and other government programs and sources.

1. Earnings Last Month

Item K3 asked respondents how much they earned last month before taxes and deductions. To avoid asking this of respondents who had already indicated in section B that they had never worked or who had indicated in section C that they were not working in the month before the interview, a series of checks were programmed to skip this item for respondents not working last month. However, these checks were based on B24 (currently working) and B36 (ever worked) only. They did not take into account respondents who indicated in B30 (worked in 2003) or in B22 (worked when limited) that they had worked. These respondents should have been asked item K2A, "Did you work last month?" but instead inappropriately skipped this item, K3 (how much earned last month), and K3a (how much was left after taxes and deductions). This programming error affected 3,274 cases. Cases with this problem are coded as .M (missing due to error) on item K2a. They are coded as .L (logical skip) on K3 and K4.

Soft edit checks were built into the instrument to flag high and low values. However, these were set to accept a wide range of responses. Based on the distribution of responses, we examined extremely low (less than $50 per month) and high values (over $5,000 per month) for both pretax and take-home pay. In most cases, we were able to evaluate the values in the context of the job-specific information provided in section C. This included considering the number of jobs the sample person currently had, the number of hours worked, the sample person's occupation, and whether the sample person was in a supported employment setting. All cases reporting less than $50 a month were sheltered workshop cases or cases in which a self-employment activity could explain low monthly wages. In most cases where $0 income was reported, sample persons were self-employed or were employed in seasonal work. Thus, none of these cases were edited during data processing. In one case with an extremely high value, it appeared that an additional "0" had been entered. This case was corrected.

We also examined differences between the pretax (K3) and post-tax pay (K3a) amounts. Some respondents reported differences in pretax (K3) and post-tax pay (K3a) that was greater than would typically be expected. For cases with differentials greater than 100 percent, the pretax or post-tax monthly income was reset to "Don't Know" (55 cases) with the exception of four cases in which it seemed an additional "0" had been entered by the interviewer. These cases were corrected. Cases with differentials between 30 percent and 99 percent were examined individually and data edited only if an obvious source of error could be identified. In one case K3a (post tax) was greater than K3 (pretax). K3a was set to missing (.D) for this case.

Finally, we compared the total monthly income calculated for all jobs currently held (based on section C data) and the monthly income reported in Section K. We expected to see differences in some cases, for example, cases in which sample persons had changed or just started a job, or where work was seasonal or where the sample person was self-employed. In 265 cases, the

difference in monthly income reported in section C and section K was at least 30 percent. In many cases, it was unclear why the discrepancy existed or which monthly income value was most correct. Due to this uncertainty, most values were not edited based on this comparison and some suspect values remain on the file. Based on feedback from interviewers, we suspect that in some cases, respondents interpreted the question to include income from sources other than earnings including SSA benefits. We recommend revising this item in future rounds to stress that only income from jobs should be included at K3.

Due to the problems associated with this item (for example, missing data and some potentially unreliable values), values for the constructed variable C_LstMnthPay (Last Month Pay) which is based on K3, were not imputed.

2. Income From Other Sources

Soft edit checks were built into the instrument to flag high and low values for income received from each source specified (K7a-K7_h). We examined values for cases in which the edit check had been suppressed (over $1,000 per month) and cases at the high and low ends of the distribution. In eight cases, the income reported was "1". These were assumed to be data entry errors and were set to missing (.D). In addition, one case with a value of $5 per month for veterans benefits was set to missing. High values were reviewed with analysts. Although some values exceeded the maximum benefit amounts for 2004, for example, $2,239 for veterans benefits, a decision was made to retain the values on the original items, although for the purposes of creating the imputed variable, values above the limit were not used in the calculation of the median from which the imputed values were derived. Additionally, values above $8,000 per month for K7_g (other regular sources) were not used in the calculation of the median for the imputed variables. Similarly, values associated with K7_h (other nonregular sources) and K14 (other government assistance) were reviewed but none were edited since none could be clearly

identified as data entry errors. One value of "1" on K15 was recoded to missing and one extremely high value was recoded to missing. In general, values of "0" for amounts received from other sources were not recoded.

3. Food Stamps Dollar Value

K12 asked respondents who had reported receiving food stamps last month to report the dollar value of the food stamps received. Respondents were asked to only include food stamps received by the sample person, not by other family members. The intent of the question was to include only food stamps that were received by the sample person or his/her family, not food stamps received by other members of the household. Based on feedback from interviewers, and some high values reported, we suspect that many respondents were confused by this question and included amounts received from other family members in the household. After discussing this issue with analysts, a decision was made to retain the values as reported on the file. We recommend revising this item in subsequent rounds to clarify the intent of the question.

4. Irregular SSI Income

Per SSA and the analysts' request, irregular SSI payments were included as nonregular income at K7_h (amount of income received from other sources not on a regular basis). For respondents who had not indicated receiving income from other nonregular sources, but who, according to SSA administrative records, had received irregular payments from SSA, K6_h was recoded as "yes" and the overages in benefit payments from administrative data was entered at K7_h. For cases that had already reported receiving income from other sources not on a regular basis, verbatim responses at K6_h regarding the source of the income were examined to determine if any included SSA, or SSI benefits. None of the responses for these cases suggested that SSA or SSI benefits was the source, so administrative data representing overages in benefit

71

payments was added to the amount already reported at K7_h for these cases. This recode affected 20 cases total.

5. Back-Coding Responses to "Other/Specify" Items

If respondents indicated receiving income from other sources on a regular basis (K6_g) or on a nonregular basis (K6_h), they were asked to specify the source. While additional categories could have been created during coding to cluster responses to these items, this would have involved creating additional amount variables in K7 to appropriately code how much income was received from each source. For cases reporting more than one source, it would not have been possible to parse out amounts. Therefore, "other" responses were not back-coded for these items.

6. Coding Open-Ended Items

Item K14 (type of assistance received from other government program) was asked as an open-ended question with no response options specified. Based on a review of the responses, categories were developed based on common responses (see Table IV.14). Coders then attempted to code the verbatim response into an established category. If the response did not fit into one of these categories, it remained as "other."

TABLE IV.14

RESPONSE CATEGORIES ADDED SECTION K AS A RESULT OF CODING

Item	Question Text	Response Categories Added
K14	What other assistance did {you/NAME} receive <u>last month</u>?	1=HOUSING ASSISTANCE 2=ENERGY ASSISTANCE 3=FOOD ASSISTANCE 4=OTHER

L.   SECTION L—SOCIODEMOGRAPHIC INFORMATION

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

1.   Living Situation

Respondents were asked at L11 to indicate whether they lived alone, lived with parents, guardians, spouse/partner, or other relative, lived with friends or roommates, lived in a group setting, or lived in some other living situation. They were then asked at L12 to describe for the place they lived. A soft edit check was built into the instrument to prompt interviewers to clarify responses in which the respondent indicated that they lived alone at L11, but also that they lived in a group setting at L12: for example a supervised apartment, group home, halfway house, personal care or board and care home, assisted living facility, nursing or convalescent home, center for independent living, or some other type of supervised group residence or facility. In some cases, this edit check was suppressed (99 cases) and the inconsistency remained. For these cases L11 was recoded to 4 ("live in another group setting").

2.   Number of Children

L17 asked how many children under age 18 lived in the sample person's household. Respondents reporting children were then asked how many of the children were their own (L19). In 49 cases, the number of own children L19 was greater than the number of children living in the household (L17). For these cases, L19 was set to missing (.D).

3.   Reporting of Household Income

Item L23Aamt asked respondents to provide their total income in 2003, or the total combined income of their household, before taxes and other deductions. Respondents for whom

it was difficult to calculate an annual amount could report their income in monthly, twice a month, weekly, bi-weekly, or daily units (recorded at L23Ahop). The amount of item nonresponse was higher for L23Aamt than other items in the survey (36 percent). Those answering "don't know" or "refused" to this item were asked to indicate which of a series of ranges described their income (L24). Of the respondents who did not respond to L23Aamt, 56 percent (1,528 cases) provided income data at L24.

The construct, C_HhInc2003, was created to combine responses provided in various units into an annual amount. We first examined high and low values on L23Aamt by unit reported (L24Ahop) and then examined high and low values on C_HhInc2003 to determine if any appeared to be invalid. There were 34 cases reporting an annual income of less than $100. With the exception of four cases, none of these respondents reported having worked in 2003. We suspect that some respondents may have thought we were asking for earnings income only rather than total income from all sources. After consulting with analysts, we decided not to set these values to missing but to retain them on the file. While household income was not imputed due to this issue, a more general construct, C_FEDPOVERTYLEVEL_ CAT1 (Household Poverty Level), which was based on reported income and household size was created and imputed. Other cases were examined on a case-by-case basis by reviewing household size and work related variables in 2003. Generally, most cases reporting a household income of $200,000 or more were edited. In some cases, it appeared that an extra "0" had been entered or that a "1" had been entered at the value rather than "1" being entered at the next item to indicate unit. These cases were corrected. In other cases, we concluded that there was an error in the unit entered, which was corrected. In all, 46 cases were edited.

4. Back-Coding Responses to "Other/Specify" Items

As mentioned above, item L11 asked respondents to indicate which of a series of items best described their living situation. Responses coded as "some other living situation" were reviewed and back-coded when possible. "Other" responses to L23ahop (how often paid) were also reviewed, although most could not be back-coded into an existing category.

5. Back-Coding Field Coded Responses

Items L12 (type of place live) was an open-ended item that required interviewers to attempt to code the respondent's verbatim response into a predetermined category. Responses that were coded as "other" by the interviewer were reviewed by coders and back-coded into existing response options when possible. Responses were not coded from "other" to a nongroup setting living situation (L12=1 through 3), however, since this would have affected instrument pathing.

M. SECTION M—CLOSING INFORMATION AND OBSERVATIONS

In this section, updated contact information was collected for the sample member and telephone information was collected for up to two contact persons for individuals in the Ticket Participant Sample. The interviewer also recorded reasons a proxy or assistance was required if appropriate, and documented special circumstances.

Back-Coding Field Coded Responses. Items M8 (how first contact related to sample person), M10 (how second contact related to sample person), M13 (how assistant/proxy related to sample person), and M14 (why assistant/proxy needed) were open-ended items that required interviewers to attempt to code the respondent's verbatim response into a predetermined category. Responses that were coded as "other" by the interviewer were reviewed by coders and back-coded into existing response options when possible. In some cases, additional categories were created during coding to cluster responses (see Table IV.15).

TABLE IV.15

RESPONSE CATEGORIES ADDED TO SECTION M AS A RESULT OF CODING

Item	Question Text	Response Categories Added
M8	How is the contact related to {you/NAME}?	11=FRIEND 12=CASEWORKER/CAREGIVER/REP PAYEE 13=GIRLFRIEND/BOYFRIEND/PARTNER 14=GUARDIAN/FOSTER PARENT/STEP PARENT 15=IN-LAW
M10	How is that person related to {you/NAME}?	11=FRIEND 12=CASEWORKER/CAREGIVER/REP PAYEE 13=GIRLFRIEND/BOYFRIEND/PARTNER 14=GUARDIAN/FOSTER PARENT/STEP PARENT 15=IN-LAW
M13	How is the assistant/proxy related to {you/NAME}?	11=FRIEND 12=CASEWORKER/CAREGIVER/REP PAYEE 13=GIRLFRIEND/BOYFRIEND/PARTNER 14=GUARDIAN/FOSTER PARENT/STEP PARENT 15=IN-LAW
M14	Why was an assistant/proxy needed?	10={NAME} FAILED COGNITIVE TEST

V. CONCLUSIONS

This report has highlighted data quality issues identified during the NBS round 1 data editing and cleaning process. In summation, both programming errors and interviewer errors resulted in the loss of some survey data. Although every effort was made to avoid errors through instrument pretesting and interviewer training, the complexity of the questionnaire made identification of all possible errors extremely difficult to anticipate and detect.

In general, while survey data processing could be made more efficient to help reach the analytic goals of the survey through the introduction of stricter range checks for values that are unusually high or low, we are hesitant to employ checks that may overwhelm and frustrate the respondent by rejecting survey responses during the interview. However, adding select consistency checks could lead to fewer data problems during the implementation of the postsurvey processing procedures. The survey instrument should be reviewed from this point of view and survey items identified that might benefit from additional range or consistency checks at the time of interview. For example, consistency checks between pre- and after-tax earnings and between earnings reported section in C (current employment) and last month income reported in section K would be useful. Adding checks must also be balanced with complicating the survey instrument with programming to account for known complexities in data being collected versus addressing data complexities after the survey is completed.

Interviewer training should also be strengthened to emphasize areas of the questionnaire where data problems were identified in this round of the survey. These areas include the use of screens to mark providers as already mentioned in section G, stressing the importance of correct data entry on job specific items, probing for sufficient information on open-ended items, and avoiding suppressing edit checks without entering comments. The goal of the improved training

should be to sensitize interviewers to areas of the questionnaire that are particularly error-prone or to the survey concepts that are particularly difficult.

A review of questionnaire design decisions and pre- and postsurvey processing is also an area where improvements can be identified. Survey items that cause confusion should be reviewed and reworked as necessary. Trade-offs between asking detailed questions that only a very small portion of the sample can answer and accepting some measurement error throughout the survey instrument need to be articulated. Questions needed for the creation of key constructed analytic variables must be identified and included in the survey instrument. Given the amount of time spent specifying and checking constructed variables, we also recommend a review of these variables at each round, bearing in mind the analytic goals of the project and the utility of each constructed variable.

Finally, time permitting, additional pretesting of the survey instrument at subsequent rounds and the paths through the instrument will help identify programming and logic errors. Additional testing is always useful, but one needs to recognize that because of the many and varied paths through the questionnaire, in combination with the large number of variables, it is very unlikely that all paths can be tested and all potential errors found.

In conclusion, the NBS data file provides a rich array of data. The data cleaning, editing, and processing tasks identified a number of places in the data file where micro-level errors were obvious. The identification of the errors suggested ways in which they can be reduced and data quality improved in the future. Future administrations of this survey will greatly benefit from the lessons learned in collecting, processing, and analyzing data from the round 1 NBS.

VI. REFERENCES

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

Mitchell, S., J. Ingels, J. Young, D. Wright, E. Smith, K. Barrett, and T. Waits. "National Beneficiary Survey." Washington, DC: Mathematica Policy Research, Inc., 2004.

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

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

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