



Signed – August 27, 2004

MEMORANDUM FOR Chester E. Bowie
Chief, Demographic Surveys Division

From: Alan R. Tupek
Chief, Demographic Statistical Methods Division

Prepared by: Stephen Ash
Victimization and Expenditures Branch
Demographic Statistical Methods Division

Subject: 2004 National Long-Term Care Survey: Sampling Specifications

I. Introduction

This memorandum describes the steps needed to produce the sample for the 2004 National Long-Term Care Survey (NLTC). For NLTC this is done in two parts: we identify the longitudinal sample and select the aged-in sample. Within the aged-in sample we select a subsample for the healthy supplement. The final product of the specification is a file of all the selected sample people with two variables: control number and a reduction code.

Before we describe the steps needed to create these two files, we will provide some general background information on NLTC as it applies to this specification. This introduction is intended to be informational to people new to NLTC and can be skipped by a knowledgeable reader.

What is NLTC?

The NLTC is a longitudinal survey of people aged 65 years old and older. The survey examines trends in chronic disability, severe cognitive impairment, and mortality (Manton 2003). The survey started in 1982 and has been conducted every five years since 1984. The Center of Demographic Studies, Duke University (CDS/DU) are currently the sponsors of the survey with assistance from a grant from the National Institute on Aging.

The Cohorts of NLTC

The longitudinal sample of NLTC has several parts which we refer to as cohorts. A specific sample person's (SP's) cohort is defined for a each round and with respect to the results of the previous rounds. Here round is used in place of the term round of interviewing. We briefly describe these cohorts and how they fit together to make the sample. Because the population of interest is always changing – people turn 65 years old

or older since the most previous round – we need to include additional people in each round. This specification describes the sampling of people aged 65-69 as of April 1, 2004, which we will also refer to as cohort 6 or the aged-in sample.

One of the important domains of interest for NLTCS is the extreme aged or people 95 years old or older. Given the special interest in this domain, the survey requires that new sample people be selected in every round in order to maintain an adequate sample size for estimation. This specification describes the sampling of these people aged 95 years old or older as of April 1, 2004, which we will refer to as cohort 9.

The NLTCS identifies sample people who are disabled or institutionalized by administering a screener questionnaire. People found to be disabled or institutionalized by the screener are given a detailed interview. This detailed interview has two types: a community and an institutional interview. The institutional interview is given to people living in an institutional setting and the community interview is given to people living at home.

The following rule applies to the detailed interview with respect to future rounds: once a SP receives a detailed interview, they are in all future rounds and they will always receive a detailed interview. People who received an institutional interview are in cohort 1 in the next round and likewise people who received an community interview are in cohort 4 in the next round.

Although NLTCS is a longitudinal survey, not every SP is interviewed in every round. Of the people who were not disabled and not institutionalized in a previous round, we only include a sample of them in the next round. This specification describes the sampling of these people, also referred to as cohort 5. We add that people not selected into cohort 5 in one round are automatically in the sample of the next round. This way no SP skips two consecutive rounds. We refer to these people who skipped a round as cohort 8.

The last cohort not yet mentioned is cohort 7. This cohort includes all SPs aged 75 years old or older, not already included in cohorts 1 or 4. These people are not eligible for the sampling of cohort 5 in order to insure we have adequate sample sizes for the population of SPs aged 75 years old or older.

The Healthy Supplement and other Sub-studies

To get a complete picture of all people aged 65 years old or older, we also select a sample of people who have never been disabled or institutionalized to receive a detailed interview. This subsample, in combination with (1) the sample people who are assigned a detailed interview by the screener questionnaire and the (2) the institutional and community cohorts, allows users to make detailed interview estimates of the entire population aged 65 years old or older. We refer to this subsample as the healthy supplement and this specification describes the sample design for it.

The NLTCs also has several sub-studies that also contribute to the analysis of impairment and disability including the Informal Caregiver Survey, Venipuncture Survey and the Next-of-Kin Sub-study. At this time none of the sub-studies have any sampling. All respondents to the NLTCs will be asked to participate in all of the sub-studies. In 1999 the analogous studies were called the Blood Sample Survey and Buccal Swab Survey (Tupek 1999a).

Table 1 summarizes the samples selected in this specification – all of which contribute to the NLTCs Samples.

Table 1: Samples Selected by this Specification

Sample	Specific Sample	General Description
6	Cohort 6	A sample of SPs aged 65-69
6H	Healthy Supplement	A subsample of Cohort 6
9	Cohort 9	A sample of SPs aged 95+
9H	Healthy Supplement	A subsample of Cohort 9
5	Cohort 5	A sample of 1999 screen-outs

Ordering the Identification of Cohorts 5 and 7

One revision to the sampling that is different than Tupek (1999b) is the order in which we identify cohorts 5 and 7. In this procedure we ask that cohort 7 be identified and then cohort 5 be sampled. In Tupek (1999b) cohort 5 was first sampled then cohort 7 identified. The two procedures are equivalent with respect to SPs age 75 years old and older, since they are in the sample with certainty. The procedures are different for those aged less than 75 years old – the 2004 has a fixed sample size sample design and the 1999 has a random sample size for SPs. The 2004 procedure is therefore better since samples from fixed sample size sample designs have smaller design variances than those with random sample sizes.

People Deceased after the Reference Date

Given that the reference date for NLTCs is April 1, 2004 and we know that we will not be able to begin interviewing prior to November 2004, the aged-in universes of interest for cohorts 6 and 9 will include deceased people. Since they are in the universe of interest they should be accounted for in the sample weighting. We plan to select a sample of them at the same rate as the regular sample. We will not send them out for interviewing, but will use them when calculating the sample weights. Given that they are non-interviews we will account for them within the non-interview adjustment or within a separate

adjustment. Note that all of the deceased SPs described will be ineligible in the next wave of the survey.

Limitation of the Sample Design

A result of sample design and the sampling frame is that NLTCs does not include people who immigrate and are aged 70 years or older. SPs aged 65 to 69 come into the sample design – both residents and immigrants, but after that no additional sampling is done for subsequent immigration. We have chosen to not complete any special procedures because this population should be small and additional cost of including them is prohibitive.

II. Files

We now describe the three input and two output files required for this specification.

A. Input Files

1. Universe Files

The Universe Files includes all living Medicare enrollees of the age groups outlined in Table 2.

Table 2: Description of Universe Files

Requested Extract	...includes SPs who were alive on April 1, 2004 and who turned...
Aged-in Universe File	65 between April 1, 1999 and March 31, 2004
95+ Universe File	95 years old or older as of April 1, 1999

Instructions for the creation of these files are given in Bowie (2004). Attachment A provides a list of the variables on the universe files.

The universe files were derived from the Enrollment Database as described in Bowie (2003). Since the Enrollment database is such an important source, we now provide further background.

The Center for Medicare and Medicaid Services (CMS) maintains the Enrollment Database is an administrative record for maintaining data on payment status, demographic characteristics, and other information necessary to administer the Medicare program.

Reports on the file are sorted by claim account number (CAN), which is the Social Security number of the primary beneficiary. All associated beneficiaries of the primary beneficiary are listed with the same CAN, immediately following the primary beneficiary's record and sorted by beneficiary identification code (BIC).

We request files from the enrollment database because we consider it a frame for the population of SPs in the U.S. that are 65 and older. We therefore can use information from the enrollment database to select samples that are representative of the U.S. population that is 65 and older and to update the information we have that is associated with SPs from previous NLTCS samples, e.g., address and date of death.

2. Longitudinal Sample File

This file contains all of the 1994 and 1999 longitudinal SPs who are still alive. We request the 1994 SPs, so that we include those SPs eligible but not selected into cohort 5 in 1999 and who thereby will comprise the 2004 cohort 8. We will assume that all SPs will have been updated using the updated address and vital statistics information requested by Bowie (2003), i.e., we have update-to-date information on the date-of-death.

For those instances where a longitudinal SP was reported deceased in the 1999 NLTCS, but CMS has no record of a Date of Death (DOD), include the SP in the longitudinal sample file. Since CDS receives monthly updates from CMS on the DOD, they would like to keep the sample consistent with the CMS update.

B. Output File

The three following files will be created by the sample selection procedure.

1. Cohort 5 Universe File

To make processing easier, we will create this subset of the Longitudinal Sample File. This file will include all SPs in the cohort 5 universe for 2004. Section III.B. describes how to create this file.

2. NLTCS Sample File

The NLTCS Sample File will include the following variables for all SPs selected into the NLTCS. Do not include the selected deceased SPs.

- ▶ Control number

- ▶ Reduction group code
- ▶ Healthy supplement date

3. Updated Sample File

Create a file of all SPs that could be expected to be in the 2004 cross-sectional sample. Specifically this file contains the following SPs:

- ▶ 2004 SPs – aged-in and alive.
- ▶ 2004 SPs – aged-in and deceased or having no address.
- ▶ 2004 SPs – longitudinal.
- ▶ 1999 SPs who screened-out and were not selected for 2004 cohort 5.
- ▶ 1999 SPs who are deceased since 1999.
- ▶ 1994 SPs who were not selected for 1999 cohort 5 and are deceased.

For each person record, provide the following variables:

- ▶ Control number.
- ▶ ID Number
- ▶ Date of birth.
- ▶ Date of death.
- ▶ 1999 cohort number.
- ▶ 2004 cohort number.
- ▶ FIPS state and county code.
- ▶ 2004 Outcome code

The second and third group will not have a cohort number for 2004 and the last group will not have a cohort number for both 1999 and 2004.

III. Preparation of the Longitudinal Sample

We now describe two steps needed for the preparation of the longitudinal sample.

A. Removal of Specific SPs

Use Table 3 to identify all those longitudinal SPs that should not be included in the final sample file used by FLD.

Table 3: Exclusions from the Longitudinal Sample

	Description	Outcome 2004
Ineligible	Deceased Prior to April 1, 2004	150
	Moved out of the country	151
Non-interview	Deceased After April 1, 2004	150
	Removal from Sample Requested by SP	118
	No Address – Unable to Locate	114

For each longitudinal SP excluded from interviewing, assign the outcome of the last column of Table 3 to identify the reason for excluding.

In sample weighting we intend to adjustment SPs with completed interviews for the non-interviews identified in Table 3.

B. Coding the 2004 Cohorts of Longitudinal SPs

Identify the 2004 cohort for all SPs in the longitudinal sample file using steps 1 and 2 below. Once a SP is determined to be in a cohort by step 1, do not consider them for step 2. Also, the value of the cohort number defines the 6th digit of the control number.

Step 1: Define Cohort 8 – Identify all living longitudinal SPs who were not in the 1999 sample, and screened out in 1994. More specifically, these SPs were in the 1994 sample in 1994 cohorts 0, 5, 6, 7, or 9 and were not on the 1999 sample. Include all of these SPs in Cohort 8 and include them in the NLTCs Sample File.

Assign a screener code of ‘S’ to all SPs of Cohort 8.

Step 2: Define Cohorts 1, 4, 7 and the Universe of Cohort 5 – With the remaining living longitudinal SPs, use Table 4 to identify their associated cohort.

Table 4: Defining Longitudinal Cohorts

If the following is true for a 1999 SP...				... then assign the following codes...	
1999 SURVEY =	SIC4 = 1 or 4	In the healthy supplement?	Aged 75 + ?	Digit 2: Screener Code	Digit 7: 2004 Cohort...
4				'N'	4
2 or 3				'N'	1
6				'N'	1
1 or 5	Yes			'N'	1
1 or 5	No	Yes		'N'	1
1 or 5	No	No	Yes	'S'	7
1 or 5	No	No	No	'S'	Include in Cohort 5 Universe File

The second column of Table 4 is identifying SPs who were selected into the healthy supplement and were a 1999 screener or detailed non-interview.

Include all of the SPs assigned to Cohorts 1, 4 and 7 in the NLTCs Sample File.

Table 5 shows how the 1999 variable SURVEY is defined.

Table 5: Values of 1999 SURVEY

Variable	Definition
1	Screener Only (Screened Out)
2	Community Survey
3	Community with Caregiver
4	Institutional
5	Screener Non-Interview
6	Detailed Non-Interview

III. Preliminary Sampling Processing

Prior to actually selecting the different samples, we need to complete some preliminary processing and get some results that we will use during sample selection.

A. Identify Deceased People on the Aged-in Universe

For all people that are aged 65 years old or older on the aged-in file and 95 years old or older on the 95+ file we receive from CMS and have a DOD, complete the action described by Table 6.

Table 6: Actions for Deceased Persons

DOD	We consider them as...	...do the following...
Prior to April 1, 2004	Ineligible	<ul style="list-style-type: none"> ▶ Do not include them on any sample file. ▶ Assign an outcome code of 450 (deceased).
On or after April 1, 2004	Eligible, Non-interview	<ul style="list-style-type: none"> ▶ Do not include them on the sample file which goes to FLD. ▶ Include them on the Updated Sample File. ▶ Assign an outcome code of 450 (deceased).

B. Identify People with No Address

Identify all people on the universe files as having no address. Although we cannot send these cases out for interviewing because we have no way of locating them, they are eligible in the survey. We will select a sample of them and adjustment for them in the weighting. Assign an outcome code of 414 (unable to locate) to each selected sample person.

C. Coding the Deceased and No Address Sort Variable

Use Table 7 to assign a value to the Deceased and No Address sort variable to all people on the aged-in universe files.

Table 7: Coding the Deceased and No Address Sort Variable

Value	Description
1	Deceased on or after April 1, 2004 Described in section III.A.

Table 7: Coding the Deceased and No Address Sort Variable

Value	Description
2	No Address Described in section III.B.
9	Otherwise

D. Coding Age Groups

Assign an Age Group code to each SP on each of the Universe Files. Attachment B specifies the values of the Age Group code.

E. Coding Stratum

The variable stratum represents the whether the original reason for applying for Medicare was because of age or some other reason. We code stratum from the CMS variable beneficiary entitlement reason code (BENE_ENTLMT_RSN_CD) originally on the Enrollment Database (CMS 1999). Table 8 specifies the coding of the variable stratum.

Table 8: Coding of Stratum

Value of STRATUM	...if the value for Original Reason for Entitlement equals...
1	0 = Beneficiary insured due to age (OASI). blank or missing value
2	Otherwise, which includes the following values: 1 = Beneficiary insured due to disability. 2 = Beneficiary insured due to end stage renal disease (ESRD). 3 = Beneficiary insured due to disability and ESRD.

F. Unduplicating Longitudinal and Aged-in SPs

Remove all longitudinal SPs from the universe files. This will insure that no SP is unnecessarily selected twice. We make no adjustment to the selection probabilities for this step because the number of SPs is very small compared with the universe totals.

G. Producing Universe Totals Needed for Sampling

Several universe totals are needed for sample selection. Table 9 defines the totals and the files from which each are derived. Each total is simply the number of records on the file of interest.

Table 9: Universe Files by Age

Total of PSU i	...is calculated from...	...and is used to calculate the Estimated National Total as ...
$N_{6,i}$	Universe File 65-69	$\hat{N}_6 = \sum_{s_1} N_{6,i} / \pi_i$
$N_{9,i}$	Universe File 95+	$\hat{N}_9 = \sum_{s_1} N_{9,i} / \pi_i$
$N_{5,i}$	Cohort 5 Universe File	$\hat{N}_5 = \sum_{s_1} N_{5,i}$

In Table 9, the third column uses the totals of the first column to calculate the estimated the national number of SPs within each universe. Here π_i is the first phase selection probabilities, which are given in Bowie (2003).

H. Calculate Sample Sizes and Sampling Fractions for NLTCS Cohorts 5, 6 and 9 by PSU.

Use the following steps of Table 10 to calculate the sample sizes and sampling fractions for each sample PSU.

Table 10: Specified Sample Sizes for NLTCs

Sample	Sample Size	Step 1	Step 2	Step 3
		National Sampling Interval	PSU Sample Size Allocation	PSU Sampling Interval
6	$n_6 = 5,600$	$TE_6 = \hat{N}_6 / n_6$	$n_{6,i} = \frac{N_{6,i}}{\pi_i TE_6}$	$TE_{6,i} = N_{6,i} / n_{6,i}$
6H	$n_{6H} = 1,000$	$TE_{6H} = n_6 / n_{6H}$		
9	$n_9 = 1,000$	$TE_9 = \hat{N}_9 / n_9$	$n_{9,i} = \frac{N_{9,i}}{\pi_i TE_9}$	$TE_{9,i} = N_{9,i} / n_{9,i}$
9H	$n_{9H} = 300$	$TE_{9H} = n_9 / n_{9H}$		
5	$n_5 = 2,300$	$TE_5 = \hat{N}_5 / n_5$		

If any of the sampling sizes of Step 2 in Table 10 are greater than the universe total for the same PSU, set the sampling interval of Step 3 equal to 1.0. This means that number of SPs allocated to the PSU is larger than the number of SPs actually in the PSU. We set the sampling fraction equal to 1.0 so that we select all of the SPs in the PSU with certainty.

Notify DSMD if any of the following errors occur:

(1) $n_{6,i} > N_{6,i}$

(2) $n_{9,i} > N_{9,i}$

IV. Sample Selection

Select the five different samples using the detailed instructions in Attachment C. The instructions for sampling use many of the same terms for each of the samples. Table 11 defines the general variables used in the sample selection.

Table 11: Variables used in Sample Selection

Variable	Definition	Value is defined as
n	The sample size, i.e., the number of SPs we are selecting.	See Table 6
N	The size of the universe.	See Tables 5 & 6
TE	The sampling fraction or take-every.	See Table 6
RN	A random number, where $0 < RN \leq 1$ and each of the values of RN are equally likely.	See Table 6
RS	The random start, having values $0 < RS \leq TE$.	$RN \times TE$
CUMNO	Identifies the units and the ordering of the universe. Valid values include the integers 1, 2, ..., N .	See Attachment C
SELECT	Identifies the sample units as a sub-sequence of equidistant integers within the values of CUMNO.	See Attachment C
q or r	Counters for the units of the sample. Valid values include the integers 0, 1, ..., $n - 1$.	See Attachment C

The variables defined in Table 11 will often be indexed. The indexing will identify the cohort number with one of the values listed in the first column of Table 11. The indexing may also include an i for the PSU or ij for the PSU and STRATUM combination. If the variable is not indexed by either i or ij then the variable applies to all SPs of interest. For example TE_{9H} is the sampling interval for all SPs in the 9H universe.

The sampling algorithm for cohorts 6 and 9 include a specialized calculation for the random start. The random start is calculated in two different ways to account for the instance when at least one SP is selected in the previous stratum or no SP is selected in the previous stratum. The specialized coding assumes that only stratum with $j = 2$ (original reason for entitlement is disability) can have no SP selected.

Table 12 describes what should be done with the units of each selected sample.

Table 12: Samples Selected by this Specification

Sample	Selected into...	
6	Cohort 6	<ul style="list-style-type: none"> ▶ Include the selected SPs in the Cohort 6 sample file ▶ Assign a '6' to the 6th digit of their control number
6H	Healthy Supplement	<ul style="list-style-type: none"> ▶ Assign a 'H' to the 14th digit of their control number
9	Cohort 9	<ul style="list-style-type: none"> ▶ Include the selected SPs in the Cohort 9 sample file ▶ Assign a '9' to the 6th digit of their control number
9H	Healthy Supplement	<ul style="list-style-type: none"> ▶ Assign a 'H' to the 14th digit of their control number
5	Cohort 5	<ul style="list-style-type: none"> ▶ Assign a '5' to the 6th digit of their control number

V. Code Assignments

Four codes need to be assigned to each SP. Two of these codes are part of the control number.

- A. 1970 Current Survey (CS) Primary Sample Unit (PSU) number – digits 7-9 of the control number.

Assign the PSU code consistent with Attachment C of Bowie (2003).

- B. Sequence number – digits 10-13 of the control number.

To assign the sequence number complete the following two steps:

1. Sort the sample by sort order A of Attachment D.
2. Consecutively number SPs starting with 0001 within each PSU.

Note that the sequence number has leading zeros in place of blanks.

- C. Healthy Supplement code – digits 14 of the control number.

We assign a value of 'H' to all SPs that were ever selected into the healthy supplement. All SPs that were never selected into the healthy supplement are assigned a value of 'O'.

Table 13 explains how to assign the healthy supplement code.

Table 13: Assignment of the Healthy Supplement Code

Cohort	Was the SP selected into 6H or 9H by this specification?	Is the 14th digit of the SP's 1999 control number a 'H'?	Assign a value of...
6 & 9	Yes		H
	No		O
1, 4, 5, 7, 8		Yes	H
		No	O

D. Reduction Group code

Complete the following two steps to assign the reduction group code.

Step 1: Sort the person records on the Sample file with sort order B of Attachment D.

Step 2: Assign a three digit code to each SP, beginning with beginning with the first SP and continuing as 000, 001, 002, ..., 999, 000, 001, ...

We assign reduction group codes in the unlikely case that the sample needs to be reduced because of budgetary problems.

E. Healthy Supplement Date

The healthy supplement date indicates which year and cohort a SP was selected into the healthy supplement. It can be used to determine the appropriate sampling interval.

Table 14 explains how to assign the values of the healthy supplement date.

Table 14: Coding of the Healthy Supplement Date

Assign the following value...	...when the following is true...	...which can be identified from the 1999 control number by...	
		Digit	...having a value of...
1994-I	1994 – initial year of the healthy supplement	13 th	H
1999-6	SIC4 = 6	5 th	6
	1999 healthy supplement	14 th	H
1999-8	SIC4 = 8	5 th	8
	1999 healthy supplement	14 th	H
1999-9	no healthy supplement selected	n/a	n/a
2004-6	the SP was selected into the 6H sample in section IV.	n/a	n/a
2004-9	the SP was selected into the 9H sample in section IV.	n/a	n/a

The first four digits of the variable denote the year when the SP was selected into the healthy supplement. The last digit identifies if it was selected in aged-in sample (cohort 6) or the 95+ supplement (cohort 9). The ‘I’ in the 1994 date denotes that all SPs were selected for the healthy supplement at that time – the initiation of the healthy supplement.

This variable replaces the need to include multiple H’s in the control number as was done in 1999 (Murphy 1999).

VI. Control Number

Assign a control number to all SP records. Attachment E summarizes the assignment of all of the digits of the control number.

The assignment of the control number as described by this specification is consistent with Ash (2003).

VIII. Verification

For quality assurance, provide the following output for verification.

A. File Requests

- ▶ Make the universe files for PSUs 104 (Philadelphia, PA) and 456 (Yavapai, AZ) available.
- ▶ Make the all of the sample files available.
- ▶ A file with both the 1999 and 2004 control numbers for the longitudinal sample that can be used to link 1999 and 2004 files.

B. Tables

Provide tables with the following:

- ▶ Values of the national sampling interval for each cohort as defined by the 3rd column of Table 8.
- ▶ Summary of the sampling for each sample group as described in Attachment F.
- ▶ Frequencies of all possible combinations of digits 2-6 of the control number, i.e., all possible combinations of the cohort numbers.

Provide the verification output to Stephen Ash, the Demographic Statistical Methods Division contact person for this specification. He is in Room 3727, Building 3 and can be reached by telephone at (301) 763 - 4294.

References

- Ash, S. (2003). Census Bureau Memorandum, "2004 National Long-Term Care Survey: Documentation of the Control Number," from Stephen Ash for Documentation, dated DRAFT.
- Center for Demographic Studies (2002). Correspondence, "Calculation of Sample Component Sizes for 2004 NLTCs."
- Center for Medicare and Medicare Services (1999). Database Dictionary, June 1, 1999.
- Bowie, C.E. (2004). Letter with subject " Requesting Files for the 2004 National Long Term Care Survey," from Chester E. Bowie to Robyn Thomas of the Center for Medicare and Medicaid Services, dated June 16, 2004.

Manton, K.G. (2002). Grant Application for “Functional and Health Changes of the Elderly,” to the Department of Health and Human Services, Public Health Services, dated June 17, 2002.

Murphy, P. (1999). Census Bureau Memorandum, “Control Numbers for the 1999 Long Term Care Survey,” from Padraic Murphy to Documentation, dated DRAFT.

Tupek, A. R. (1999a). Census Bureau Memorandum, “REVISED Sample Design and Sampling Specifications for Blood Sample Survey and Buccal Swab Survey,” from Alan R. Tupek to Chester E. Bowie, DRAFT.

Tupek, A. R. (1999b). Census Bureau Memorandum, “Sampling Specifications for the 1999 Long-Term Care Survey (LTC): Longitudinal Portion, and Longitudinal Portion of the Healthy Supplement and the 95+ Supplement,” from Alan R. Tupek to Chester E. Bowie, dated May 18, 1999.

Attachments

cc:

R. Dopkowski (DSD)
E. Williams
P. Flanagan (DSMD)
K. King
S. Ash
K. Manton (CDS/DU)
R. Kulka
N. Holt
R. Schwartz

Table A1: File Format for Requested Files

Relation Description	Element Descriptions	Element Field Name
Beneficiary Identification	Beneficiary Claim Account Number (CAN) (SSA account no.)	BENE_CLM_ACNT_NUM
	Beneficiary Identification Code (BIC)	BENE_IDENT_CD
	Beneficiary Birth Date	BENE_BIRTH_DT
	Beneficiary Death Date	BENE_DEATH_DT
	Beneficiary Sex Identification Code	BENE_SEX_IDENT_CD
	Beneficiary Race Code	BENE_RACE_CD
	Beneficiary First Name	BENE_GVN_NAME
	Beneficiary Middle Name	BENE_MDL_NAME
	Beneficiary Surname	BENE_SRNM_NAME
	Beneficiary Mailing Contact Address	BENE_MLG_CNTCT_ADR
	Beneficiary Representative Payee Switch	BENE_RPRSNTV_PYE_SW
Social Security Numbers	Beneficiary Social Security Number	BENE_SSN_NUM
Entitlement Reason	Beneficiary Entitlement Reason Code (original reason for entitlement)	BENE_ENTLMT_RSN_CD
Residence	Beneficiary Mailing Contact Zip Code	BENE_MLG_CNTCT_ZIP_CD
	Beneficiary Residence (SSA Standard State Code) (obtain for current and any prior address)	BENE_RSDNC_SSA_STD_STATE_CD
	Beneficiary Residence (SSA Standard County Code) (obtain for current and any prior address)	BENE_RSDNC_SSA_STD_CNTY_CD

Defining the Variable Age Group

Table B1 defines the values of the variable Age Group.

Table B1: Definition of the Variable Age Group

Age Group Value	Age as if April 1 st , 2004
1	65 - 69
2	70 - 74
3	75 - 79
4	80 - 84
5	85 - 89
6	90 - 94
7	95 - 99
8	100 - 104
9	105 - 109
10	110 - 114
11	115 - 119
12	120 - 124
13	125 - 129

Sampling of Cohort 6 (Aged-in)

Complete the following three steps to select the Aged-in sample.

Step 1: Sort the person records on the Aged-in Universe file with sort order A of Attachment D.

Step 2: Initialize starting values.

- ▶ Initialize $RN_6 = 0.471648$. The variable RN is the random number that defines the systematic sample.

Note that i indexes 143 PSUs and j indexes the two STRATUM.

Step 3: Starting with the first record, complete steps 3a through 3e for all records, maintaining the order created by step 1.

Step 3a: For every “new” PSU and STRATUM combination ij , i.e., the first SP of every PSU or STRATUM, complete the following:

- ▶ Initialize the random start for all PSU, except the first with the following rule:

▶ If ($j = 1$ and $n_{i-1,2} = 0$) then

$$RS_{6,i+1,1} = TE_{6,i+1} \left[1 - \frac{N_{6,i,1} - \left[RS_{6,i,1} + (n_{6,i,1} - 1) \times TE_{6,i} \right] + N_{i,2}}{TE_{6,i}} \right].$$

▶ If ($j = 2$) or ($j = 1$ and $n_{i-1,2} > 0$) then

$$RS_{6,i+1,j+1} = TE_{6,i+1} \left[1 - \frac{N_{6,ij} - \left[RS_{6,ij} + (n_{6,ij} - 1) \times TE_{6,ij} \right]}{TE_{6,ij}} \right].$$

- ▶ Initialize $q_{ij} = 0$. The variable q_{ij} is the counter for sampling of cohort 6 and has values 0, 1, ..., $n_{6,ij} - 1$.
- ▶ Initialize $CUMNO_{6,ij} = 0$. The variable $CUMNO_{6,ij}$ is the cumulative number for the cohort 6 universe and has values 0, 1, 2, ..., $N_{6,ij}$.

Sampling of Cohort 6 (Aged-in)

Step 3b: Increment $CUMNO_{6,ij}$ by one unit, i.e., $CUMNO_{6,ij} = CUMNO_{6,ij} + 1$.

Step 3c: Calculate $SELECT_{6,ij} = \lceil RS_{6,ij} + (q_{ij} \times TE_{6,i}) \rceil$.

where $\lceil - \rceil$ represents the ceiling function, i.e., it returns the next greatest integer, and the value of $TE_{6,i}$ is defined in Table 7.

Step 3d: If $SELECT_{6,ij} = CUMNO_{6,ij}$ for a specific SP, do the following:

- ▶ Select the SP into cohort 6.
- ▶ Increment q_{ij} by one unit, i.e., $q_{ij} = q_{ij} + 1$.

Sampling for the Cohort 6 Healthy Supplement

Complete the following three steps to select the sample of 1999 Screen-outs, i.e., cohort 5.

Step 1: Sort the SP records on the Cohort 6 healthy supplement universe, i.e., the cohort 6 selected sample, with sort order A of Attachment D.

Step 2: Initialize starting values.

- ▶ Initialize $RN_{6H} = 0.1586326$. The variable RN is the random number that defines the systematic sample.
- ▶ Initialize $q = 0$. This is the counter for the sampling of 6H and has values 0, 1, ..., $n_{6H} - 1$.
- ▶ Initialize $CUMNO_{6H} = 0$. The variable $CUMNO_{6H}$ is the cumulative number for 6H universe and has values 1, 2, ..., N_{6H} .

Step 3: Calculate the random start for the first SP, i.e., set $RS_{6H} = TE_{6H} \times RN_{6H}$.

Step 4: Increment $CUMNO_{6H}$ by one unit, i.e., $CUMNO_{6H} = CUMNO_{6H} + 1$.

Step 5: Calculate $SELECT_{6H} = \lceil RS_{6H} + (q \times TE_{6H}) \rceil$.

where $\lceil - \rceil$ represents the ceiling function, i.e., it returns the next greatest integer, and the value of TE_{6H} is defined in Table 7.

Step 6: If $SELECT_{6H} = CUMNO_{6H}$ for a specific SP, do the following:

- ▶ Select the SP into the healthy supplement for cohort 6 – assign a value of ‘H’ to the 14th digit of their control number.
- ▶ Increment q by one unit, i.e., $q = q + 1$.

Sampling for Cohort 9 (95+ Supplement)

Complete the following three steps to select the 95 + Supplementary sample.

Step 1: Sort the SP records on the 95+ Universe file with sort order A of Attachment D.

Step 2: Initialize starting values.

- ▶ Initialize $RN_9 = 0.517936$. The variable RN is the random number that defines the systematic sample.
- ▶ For the first SP of the first PSU and STRATUM, set $RS_{9,ij} = TE_{9,i} \times RN_9$.

Note that i indexes 143 PSUs and j indexes the two STRATUM.

Step 3: Starting with the first record, complete steps 3a through 3e for all records, maintaining the order created by step 1.

Step 3a: For every “new” PSU and STRATUM combination ij , i.e., the first SP of every PSU or STRATUM, complete the following:

- ▶ Initialize the random start for all PSU, except the first with the following rule:

▶ If ($j = 1$ and $n_{i-1,2} = 0$) then

$$RS_{9,i+1,1} = TE_{9,i+1} \left[1 - \frac{N_{9,i,1} - [RS_{9,i,1} + (n_{9,i,1} - 1) \times TE_{9,i}] + N_{i,2}}{TE_{9,i}} \right].$$

▶ If ($j = 2$) or ($j = 1$ and $n_{i-1,2} > 0$) then

$$RS_{9,i+1,j+1} = TE_{9,i+1} \left[1 - \frac{N_{9,ij} - [RS_{9,ij} + (n_{9,ij} - 1) \times TE_{9,ij}]}{TE_{9,ij}} \right].$$

- ▶ Initialize $q_{new} = 0$. The variable q_{ij} is the counter for the sampling of cohort 9 and has values 0, 1, ..., $n_{9,ij} - 1$.
- ▶ Initialize $CUMNO_{9,new} = 0$. The variable $CUMNO_{9,ij}$ is the cumulative number for the cohort 9 universe and has values 0, 1, 2, ..., $N_{9,ij}$.

Sampling for Cohort 9 (95+ Supplement)

Step 3b: Increment $CUMNO_{9,ij}$ by one unit, i.e., $CUMNO_{9,ij} = CUMNO_{9,ij} + 1$.

Step 3c: Calculate $SELECT_{9,ij} = \lceil RS_{9,ij} + (q_{ij} \times TE_{9,i}) \rceil$.

where $\lceil - \rceil$ represents the ceiling function, i.e., it returns the next greatest integer, and the value of $TE_{9,i}$ is defined in Table 7.

Step 3d: If $SELECT_{9,ij} = CUMNO_{9,ij}$ for a specific SP, do the following:

- ▶ Select the SP into cohort 9.
- ▶ Increment q_{ij} by one unit, i.e., $q_{ij} = q_{ij} + 1$.

Sampling for the Cohort 9 Healthy Supplement

Complete the following three steps to select the sample healthy supplement for cohort 9.

Step 1: Sort the SP records on the cohort 9 healthy supplement universe, i.e., the cohort 9 selected sample with sort order A of Attachment D.

Step 2: Initialize starting values.

- ▶ Initialize $RN_{9H} = 0.972541$. The variable RN is the random number that defines the systematic sample.
- ▶ Initialize $q = 0$. This is the counter for the sampling of cohort 5 and has values 0, 1, ..., $n_{9H} - 1$.
- ▶ Initialize $CUMNO_{9H} = 0$. The variable $CUMNO_{9H}$ is the cumulative number for 9H universe and has values 1, 2, ..., N_{9H} .

Step 3: Calculate the random start for the first SP, i.e., set $RS_{9H} = TE_{9H} \times RN_{9H}$.

Step 4: Increment $CUMNO_{9H}$ by one unit, i.e., $CUMNO_{9H} = CUMNO_{9H} + 1$.

Step 5: Calculate $SELECT_{9H} = \lceil RS_{9H} + (q \times TE_{9H}) \rceil$.

where $\lceil - \rceil$ represents the ceiling function, i.e., it returns the next greatest integer, and the value of TE_{9H} is defined in Table 7.

Step 6: If $SELECT_{9H} = CUMNO_{9H}$ for a specific SP, do the following:

- ▶ Select the SP into the healthy supplement for cohort 9 – assign a value of ‘H’ to the 14th digit of their control number.
- ▶ Increment q by one unit, i.e., $q = q + 1$.

Sampling for Cohort 5 – Sample of 1999 Screen-outs

Complete the following three steps to select the sample of 1999 Screen-outs, i.e., cohort 5.

Step 1: Sort the SP records on the Cohort 5 Universe file with sort order A of Attachment D.

Step 2: Initialize starting values.

- ▶ Initialize $RN_5 = 0.823374$. The variable RN is the random number that defines the systematic sample.
- ▶ Initialize $q = 0$. This is the counter for the sampling of cohort 5 and has values 0, 1, ..., $n_5 - 1$.
- ▶ Initialize $CUMNO_5 = 0$. The variable $CUMNO_5$ is the cumulative number for cohort 5 universe and has values 1, 2, ..., N_5 .

Step 3: Calculate the random start for the first SP, i.e., set $RS_{5,i} = TE_5 \times RN_5$.

Step 4: Increment $CUMNO_5$ by one unit, i.e., $CUMNO_5 = CUMNO_5 + 1$.

Step 5: Calculate $SELECT_5 = \lceil RS_5 + (q \times TE_5) \rceil$.

where $\lceil - \rceil$ represents the ceiling function, i.e., it returns the next greatest integer, and the value of TE_5 is defined in Table 7.

Step 6: If $SELECT_5 = CUMNO_5$ for a specific SP, do the following:

- ▶ Select the SP into cohort 5 – include them in the Sample File and assign a value of ‘5’ to the 7th digit of their control number.
- ▶ Increment q by one unit, i.e., $q = q + 1$.

Table D1: Sort Orders

Sort Order	Specification of Exact Sort
A	<ul style="list-style-type: none"> ▶ LTC PSU ▶ STRATUM ▶ State and county ▶ Deceased and No Address Variable ▶ Age Group ▶ Race ▶ Sex ▶ Age ▶ Last name ▶ First name ▶ Middle name
B	<ul style="list-style-type: none"> ▶ LTC PSU ▶ STRATUM ▶ State and county ▶ Cohort ▶ Age Group ▶ Race ▶ Sex ▶ Age ▶ Last name ▶ First name ▶ Middle name

Table E1: Digits of the 2004 NLTCs Control Number

Digit(s)	Description	Value of Digit(s) by 2004 Cohorts		
		6 & 9	1, 4, 5 & 7	8
1	Screener Code	‘S’	see Section IV.B.	‘S’
2	Sample Identification Code 1 – 1984	see Table 4	Digit 2 of 1999 Control Number	Digit 3 of 1994 Control Number
3	Sample Identification Code 2 – 1989	see Table 4	Digit 3 of 1999 Control Number	Digit 4 of 1994 Control Number
4	Sample Identification Code 3 – 1994	see Table 4	Digit 4 of 1999 Control Number	Digit 5 of 1994 Control Number
5	Sample Identification Code 4 – 1999	see Table 4	Digit 5 of 1999 Control Number	‘0’
6	Sample Identification Code 5 – 2004	see Table 8	see Section IV.B.	‘8’
7-9	1970 Current Survey (CS) Primary Sample Unit (PSU) number	see section V.	see section V.	see section V.
10-13	Sequential number	see section V.	see section V.	see section V.
14	Healthy supplement code	see Table 10	see Table 10	see Table 10
15	Informal Caregiver Survey Code	‘C’	‘C’	‘C’

The term “use all prior values” in Table 8 means that the values of the past cohorts for a SP do not change, so use the old values.

Output for Sampling Verification

Table F1 specifies the output for the verification of sample 6. Provide the same table and output for sample 9.

Table F1: Table Format for Sampling Output of Sample 6 and 9

PSU	π_i	$N_{6,i}$	$n_{6,i}$	$TE_{6,i}$	$wgt_{6,i}$	$chk_{6,i}$
101						
102						
...						
463						
Total		***	***			***

where

$N_{6,i}$ the universe size as specified in Table 9.

$\hat{n}_{6,i}$ the actual sample size that was selected.

π_i the first order selection probability as in Bowie (2003).

$wgt_{6,i}$ the sampling weight and is defined as $wgt_{6,i} = N_{6,i} / (\pi_i n_{6,i})$.

$chk_{6,i}$ a check of the sampling weights and is defined as the sum of the weights over the sample, i.e., $chk_{6,i} = \hat{n}_{6,i} \times TE_6$. This total should equal $\hat{N}_{6,i}$.

*** identifies a requested total for the column.

Output for Sampling Verification

Table F2 specifies the output for the verification of cohort 5. Provide the same output for 6H and 9H.

Table F2: Table Format for Sampling Output of Cohorts 5

PSU	$N_{5,i}$	$n_{5,i}$	$chk_{5,i}$
101			
102			
...			
463			
Total	***	***	***

where

n_{5i} the actual number of SPs selected into cohort 5 for PSU i .

N_{5i} the actual number of SPs eligible for sampling in cohort 5 for PSU i .

$chk_{5,i}$ a check of the sampling fraction and is defined as the actual sampling interval within a given PSU, i.e., $chk_{5,i} = N_{5i} / n_{5i}$.

*** identifies a requested total for the column.

The following two items should be consistent within the table:

- ▶ The value of $chk_{5,i}$ within each PSU should be close to the value calculated in Table 7 and used to select the cohort 5 sample.
- ▶ The totals for N_{5i} and n_{5i} should be close to the values used to calculate the TE and defined in Table 8.