

2004 NATIONAL LONG TERM CARE SURVEY

CALCULATING THE BASE WEIGHT – REVISION

JULY 12, 2006

**2004 National Long Term Care Survey
Calculating the Base Weight - Revision
July 12, 2006**

I. Introduction

This document explains how to calculate the base weights for the 2004 National Long Term Care Survey (NLTCS). The sample persons (SPs) should include all those sent for interviewing and all those non-interview SPs excluded during the sample selection for various reasons (U.S. Census 2006a; p.2). To assist with the subsequent variance estimation also assign a half sample and psuedo strata code to each SP. Assign these two codes consistent with U.S. Census (2006b).

This revision replaces Tupek (2005) and includes a corrected calculation in Table 5.

We define the base weight as

$$BW_k = SI_{1\&2,k} H SI_{5,k}$$

where

BW_k	the base weight
$SI_{1\&2,k}$	the sampling interval for the first and second stages of sampling
$SI_{5,k}$	the sampling interval for cohort 5 sampling
k	the index for people in the sample

In the rest of this memorandum we will discuss what values to assign for each of the sampling intervals that make the base weight. We also discuss revising the weights for SPs aged 95 years old or older, non-interview adjustments or first and second stage ratio adjustments.

II. Sampling Intervals

We now discuss how to calculate the base weights with respect to the first and second stage sampling intervals. We also provide instructions for defining the base weights for SPs aged 95 years old and older and for calculating the healthy supplement factor.

Sampling Intervals for the First and Second Stage Samples

For all of the waves of NLTCS the combined first and second stage sampling interval are constant across all PSUs, i.e., the sample design is self-weighting within wave. The allocation to the PSUs that ensures the constant sampling interval for 2004 is given in Tupek (2004, p.12).

Assign the base weight to each SP with 2004 outcome codes 118 and 150 or blank. A SPs exact base weight is determined by the year the SP was in an incoming cohort as listed in Table 1.

Table 1: Combined First and Second Phase Probabilities of Selection

Year Aged-in	Incoming Cohort	$SI_{1\&2,k}$	Reference
1982	SIC1 = 1, 2, 3, 4	730.5909 ¹ [= 477.415818 H(101/66)]	(Jones 1982; p.2)
	SIC1 = 5	1,610.6208 ² [= 730.5909 H(97/44)]	(Jones 1985; p.2)
	SIC1 = 0, SIC2 = 0, SIC3 = 8	See instruction below	
1984	SIC1 = 6	730.5909	(Jones 1985; p. 2)
1989	SIC2 = 6	1,505.8360	(Waite 1994c; p.7)
1994	SIC3 = 6	2,244.7064	(Waite 1994c; p.7)
	SIC3 = 9	373.7185	
1999	SIC4 = 6	1,633.1217	(Tupek 1999b; p.7)
	SIC4 = 9	688.0678	(Tupek 1999b; p.G-3)
2004	SIC5 = 6	1,617.3365 [= 9,100,752.7106 / 5,627]	(U.S. Census 2006a; p.3)
	SIC5 = 9	242.0218 [= 411,437.018452 /1,700]	(U.S. Census 2006a; p.3)

See U.S. Census (2006c) for the definitions of the cohorts. Also note that $SI_{1\&2,k}$ is equivalent to SI in the final output file.

After assigning the base weights according to Table 1, there will be SPs who still have not been assigned a base weight. These SPs can be identified as having SIC1 = 0, SIC2 = 0, and SIC3 = 8. For some unknown reason we do not have their correct SIC codes. CDS will assign the base weight to these SPs using the values of the 1999 base weights. Attachment D provides

¹ The base weight of 730.5909 appears throughout this document. It was assigned based on information available from existing documentation. This base weight is currently under review by the Center for Demographic Studies.

² The base weight of 1,610.6208 appears throughout this document. It was assigned based on information available from existing documentation. This base weight is currently under review by the Center for Demographic Studies.

an alternative method for assigning the base weight to these people that was suggested by Census.

Sampling Interval for the Cohort 5 Sampling

For each wave of NLTCS we select a subsample of SPs who have screened-out to not participate in the given wave of NLTCS. We say a SP has screened-out when they (1) are not disabled, (2) not in an institution, and (3) did not screen-in in a previous wave. The sampling of cohort 5 reduces the overall response burden and allows the sample to be allocated to the primary population of interest – people in institutions and with disabilities.

Apply the sampling interval of $SI_{5,k} = 1.81$ [= 4,163 / 2,300] as reported in U.S. Census (2006a) to all SPs in cohort 5 for 2004. Also apply a factor of 1.0 to all SPs not in cohort 5.

For historical completeness Table 2 also provide all of the previous weighting factors for cohort 5.

Table 2: Sampling Intervals for Cohort 5

Year in Cohort 5	SI_{5k}	Reference
1989	2.9638	(Waite 1989; p.2)
1994	1.9884	(Waite 1994c; p.7)
1999	1.94787 [= 3,064 / 1,573]	(Tupek 2005; p.D5)
2004	1.81 [= 4,163 / 2,300]	(U.S. Census 2006a; p.3)

Using the sampling intervals of Tables 1 and 2, a data user can generate the screener sample weight for any wave of NLTCS. To derive an estimate of a previous wave, include the cohort 5 factor of Table 2 that corresponds to the same wave in the weight of all cohort 5 SPs.

Base Weights for SPs 95+

Use Table 3 to assign the base weights to all SPs aged 95 years old and older.

Table 3: Groups for Combining the Screener Base Weight

Age	...was selected in...	...and can be identified by...	...assign a BW=...
95 to 99	1982	SIC1 = 1, 2, 3, 4	94.6923
	1982	SIC1 = 5	208.7535
	2004, Cohort 9	SIC5 = 9	210.6532
100 to 104	1982	SIC1 = 1, 2, 3, 4	65.2325
	1982	SIC1 = 5	143.8080
	1999, Cohort 9	SIC4 = 9	198.3239
	2004, Cohort 9	SIC5 = 9	150.6536
105 +	1982	SIC1 = 1, 2, 3, 4	26.3445
	1982	SIC1 = 5	58.0777
	1994, Cohort 9	SIC3 = 9	71.5812
	1999, Cohort 9	SIC4 = 9	159.2718
	2004, Cohort 9	SIC5 = 9	130.9159

The base weights of Table 3 represent a revision of the original values of Table 1. The revision is needed because each of the sets of SPs of the same incoming cohort can be used to estimate the same totals. We “combine the samples” using the methodology outlined in Attachment B. The exact calculation is given in Attachment C.

Sampling Interval for the Healthy Supplement

At the time a SP was selected into the NLTCs, an additional stage of sampling was completed for the healthy supplement. We use this subsample to select SPs who are “healthy”, i.e., SPs who do not screen-in and therefore do not receive a detailed interview with certainty.

We select the sample of the healthy supplement at the same time that we select the original sample for operational reasons. Since the healthy supplement is a subsample of people who screen-out, we could select the sample after the completing screener interviewing. However because of timing, i.e., we do not have the time to conduct screener interviews, return results,

then select the sample; we randomize the selection of the healthy supplement at the time of the original sample.

Apply the healthy supplement factor to all SPs for which the following two conditions apply:

- (1) The SP was selected into the healthy supplement.
- (2) The SP is “unimpaired,” i.e., they may have received a community interview in a previous wave, but they never received it because of disability.

A SP is in the healthy supplement if they have a ‘H’ in their 14th digit of their 2004 control number. See U.S. Census (2006) for more on the characters of the 2004 control number.

A SP is “unimpaired” in a given wave of interviewing if they were not institutional or disabled, and therefore did not receive a detailed interview. A sample person is “unimpaired” if they have been healthy in all waves in which they have been in the survey.

Table 4 summarizes how to identify a person who is “unimpaired.”

Table 4: Defining “Unimpaired” for 2004 Sample People

A SP is “unimpaired” if both of the following conditions are true:	...which is defined as...
(1) The SP was not institutionalized since 1994.	SIC3 NE 4 and SIC4 NE 4 and SIC5 NE 4
(2) The SP has not been disabled or institutionalized since 1994.	DISGRP ₁₉₉₄ NE 2, 3, 4, 5 or 6 and DISGRP ₁₉₉₉ NE 2, 3, 4, 5 or 6 and DISGRP ₂₀₀₄ NE 2, 3, 4, 5 or 6

Apply the healthy supplement using all the conditions except DISGRP₂₀₀₄ NE 2, 3, 4, 5 or 6. Then in the weighting when DISGRP₂₀₀₄ is available, we’ll change the healthy supplement factor to 1.0 for all sample people who are disabled or institutional in 2004 (DISGRP₂₀₀₄ = 2, 3, 4, 5 or 6).

The variable disability group (DISGRP) is defined in Table A1 of Attachment A. The value of 4 for SIC indicates that the SP received a detailed institutional interview in the previous wave. Both checks start with the 1994 wave since the healthy supplement started in 1994.

This definition of unimpaired insures that SPs selected into the healthy supplement and were disabled or institutionalized are assigned a healthy supplement factor of 1.0. We do not want them to get the factor greater than 1.0 because if they were once previously disabled or institutionalized, then they would have been assigned a detailed community with certainty.

For all those SPs of the healthy supplement satisfying the two conditions in Table 5, apply the sampling interval of Table 5 according to the healthy supplement date.

Table 5: Healthy Supplement Sampling Fractions

Healthy Supplement Date	Original Population	HSF_k	Reference
1994-I	All SPs	10.3605	(Waite 1994a; p.13) (Waite 1994b; p.4) (Waite 1994c; p.8)
1999-6	Cohort 6	25.0883	(Tupek 1999a; p. 4)
1999-8	Cohort 8	[= (5,500 + 1,600) / 283]	(Tupek 1999b; p.13)
2004-6	Cohort 6	5.6030 [= 5,603 / 1,000]	(U.S. Census 2005a; p.3)
2004-9	Cohort 9	5.2800 [= 1,584 / 300]	(U.S. Census 2006a; p.3)

Otherwise $HSF_k = 1.0$.

The healthy supplement date was assigned by Tupek (2004). It represents the date in which a SP was selected into the healthy supplement. Since the healthy supplement started in 1994, all SPs selected into NLTCs prior to and including 1994 were selected into the healthy supplement in 1994. For 1999 and 2004 the healthy supplement coincides with the year SP was selected into NLTCs.

Apply a factor of 1.0 to all SPs not in the healthy supplement.

III. Units not sent for Interviewing

We now discuss the assignment of the outcome code to units that were selected but not sent for interviewing and also document the reason for not being sent for interviewing.

Assignment of the 2004 Outcome Code

During sample selection we identified several other types of longitudinal SPs who we identified as being ineligible or non-interviews and thereby will not be sent out for interviewing. Assign an outcome code to each of these SPs who are on the base weight file but not sent out for interviewing. Table 6 is from U.S. Census (2005a; p. 2) and provides the values of the outcome code to assign and the counts by each of these outcomes.

Table 6: Assignment of 2004 Outcome Code on Base Weight File

Description	2004 Outcome	Number of SPs	
Non-Interview	Deceased After April 1, 2004	150	272

Table 6: Assignment of 2004 Outcome Code on Base Weight File

	Description	2004 Outcome	Number of SPs
Non-Interview	Deceased After April 1, 2004	150	272
	Removal from Sample Requested by SP	118	4
Ineligible	Deceased Prior to April 1, 2004	148	6,319
	Reported Deceased in 1994 and 1999, but no CMS DOD	149	7
	Moved out of the country	151	14
	Duplicate of previous longitudinal SP	198	14
	Not Selected for Cohort 5	199	1,864
	Sent to Interviewing		12,886

The aged-in sample also included people who were deceased after April 1, 2004. These SP will also not be sent for interviewing and should be assigned a value of 150 for the 2004 outcome code. Table 7 summarizes the counts for the aged-in sample.

Table 7: Aged-in Sample Counts

Description	2004 Outcome	Counts for Cohort...		
		6	9	
Non-interview	Deceased After April 1, 2004	150	24	116
	Sent for Interviewing	blank	5,603	1,570

All SPs sent for interviewing should have a blank value for the variable outcome, with one exception. There are 14 SPs of cohort 9 who were found to be duplicates of SPs of the longitudinal sample after sample selection. Since they were already sent for interviewing before they were found to be duplicates, we instructed Field Division to code them as ineligible. The 14 duplicates are identified on the base weight file with a final outcome code of 198.

IV. Output

The result of this memorandum is a file containing a record for each SP of the 2004 NLTC. For each SP include the following variables:

- < Control number
- < Sequence Number (referred to as either SEQ or IDNUM)
- < Base weight (BW)
- < Original Sampling Interval (SI)
- < Healthy Supplement Factor (HSF)
- < Pseudo Stratum Code (P_STRAT)
- < Half sample code (HALFSAMP)
- < Healthy supplement Code (HS_CODE)
- < Have Always Been Healthy Code (HABH)
- < Healthy supplement date (HS_DATE)
- < Sample date (S_DATE)
- < Sampled Region (REGION)
- < 1984 Sample Component (SCOMP84)
- < AGE and AGE_GROUP
- < OUTCOME2004

References

- Jones, C. (1982). "Weighting Specifications for the Long Term Care Survey (LTC)," Census Bureau Memorandum from Charles D. Jones to Thomas C. Walsh dated August 19, 1982.
- Jones, C. (1985). "1984 LTC Weighting Specifications for Cross-Sectional Estimates," Census Bureau Memorandum from Charles D. Jones to Thomas C. Walsh dated March 8, 1985.
- Tupek, A.R. (1999a). "Sampling Specifications for the 1999 Long-Term Care Survey (LTC): Longitudinal Portion and Longitudinal Portion of the Healthy Supplement, the 95+ supplement, the Blood Sample, the Buccal Swab Sample, and the Caregiver Sample," Census Bureau Memorandum for Chester E. Bowie from Alan R. Tupek, dated May 18, 1999.
- Tupek, A.R. (1999b). "Sampling Specifications for the 1999 Long-Term Care Survey: the Aged-In Cohort, the 95+ Supplement, the Aged-In Portion of the Healthy Supplement, the Blood Sample, the Buccal Swab Sample, and the Caregiver Sample," Census Bureau Memorandum for Chester E. Bowie from Alan R. Tupek, dated May 18, 1999.
- Tupek, A.R. (2004). Census Bureau Memorandum, "2004 National Long-Term Care Survey: Sampling Specifications," from Alan R. Tupek for Chester E. Bowie, dated August 27, 2004.

Tupek, A. R. (2005). "1999 National Long Term Care Survey, Weighting for Cross-Sectional Estimation," Census Bureau document from Alan R. Tupek for Kenneth G. Manton and Richard A. Kulka, dated May 16, 2005.

Tupek, A. R. (2005). "2004 National Long-Term Care Survey: Calculating the Base Weight," from Alan R. Tupek for Chester E. Bowie, dated May 16, 2005.

U.S. Bureau of the Census. (2006a) "2004 National Long Term Care Survey: Sampling Results - Revision," June 29, 2006.

U.S. Bureau of the Census. (2006b) "2004 National Long Term Care Survey: Assigning and Using the Pseudo Strata and Half Sample Codes," June 29, 2006.

U.S. Bureau of the Census. (2006c) "2004 National Long Term Care Survey: Revised Documentation of the Control Number and the Healthy Supplement Date," July 12, 2006.

Waite, P.J. (1989). "1989 LTC Weighting Specifications for Cross-Sectional Estimates," Census Bureau Memorandum from Preston J. Waite to B. Gregory Russell, dated November 6, 1989.

Waite, P.J. (1994a). "Sampling Specifications for the 1994 Long-Term Care Survey (LTC): Longitudinal Portion and Longitudinal Portion of the Healthy Supplement (Step 2)," Census Bureau Memorandum for Sherry L. Courtland from Preston J. Waite, dated May 6, 1994.

Waite, P.J. (1994b). "Sampling Specifications for the 1994 Long-Term Care Survey: the Aged-In Component, the 95+ Supplement, the Aged-In Portion of the Healthy Supplement," Census Bureau Memorandum for Sherry L. Courtland from Preston J. Waite, dated May 6, 1994.

Waite, P.J. (1994c). "REISSUE: 1994 Long Term Care (LTC) Survey Weighting Specifications for Cross-Sectional Estimates," Census Bureau Memorandum from Preston J. Waite to Sherry L. Courtland, dated November 20, 1994.

Recommended reference: U.S. Bureau of the Census (2006). "2004 National Long Term Care Survey: Calculating the Base Weights - Revision," July 12, 2006.

2004 National Long Term Care Survey

Table A1: Defining Disability for a Sample Person

A SP is ...	DISGRP =	Definition
Disabled	2	IADL Only
	3	1 - 2 ADLS
	4	3 - 4 ADLS
	5	5 - 6 ADLS
	6	Institutionalized
	Non-disabled	1
7		Non-interview – Deceased
8		Non-interview – Other

Combining Base Weights for the 95+ Samples

We need to combine the base weights of the sample persons (SPs) aged 95 years old and older (95+) since we selected their samples at different times from the same universe. Since each of the different samples of the 95+ universe could be used to estimate the same characteristics of interest, using the base weights without an adjustment will produce an overestimate. Although the samples are not strictly independent because of unduplication, we assume they are independent with one another.

Table B1 summarizes the incoming samples of SPs aged 95 years old or older by five year age groups.

Table B1: Groups for Combining the Screener Base Weight

Universe	...was selected in...	...and can be identified by...
95 to 99	1982	SIC1 = 1, 2, 3, 4
	1982	SIC1 = 5
	2004, Cohort 9	SIC5 = 9
100 to 104	1982	SIC1 = 1, 2, 3, 4
	1982	SIC1 = 5
	1999, Cohort 9	SIC4 = 9
	2004, Cohort 9	SIC5 = 9
105 +	1982	SIC1 = 1, 2, 3, 4
	1982	SIC1 = 5
	1994, Cohort 9	SIC3 = 9
	1999, Cohort 9	SIC4 = 9
	2004, Cohort 9	SIC5 = 9

We also define n_i as the number of completed interviews and non-interviews for the given group. They can be identified as having an outcome code that is blank (sent out for interviewing), 118

(Deceased After April 1, 2004) or 150 (Removal from Sample Requested by SP). See also Table 8.

To explain the factors we use the example of domain of people aged 100 to 104 years old. Here we need to find N_{82} , N_{99} and N_{04} , where $0 < N_{82}, N_{99}, N_{04} < 1$, $N_{82} + N_{99} + N_{04} = 1$ and such that

$$\hat{T}_{y,combined} = \phi_{82} \hat{T}_{y,82} + \phi_{99} \hat{T}_{y,99} + \phi_{04} \hat{T}_{y,04} \quad (B.1)$$

Here we say that $\hat{T}_{y,i}$ and N_k are the estimate of the total for a variable of interest y and the factor for combining the samples, respectively. The k is indexing the year a given incoming sample was selected.

We would like a set of factors that pool the resultant estimator of the different samples together so as to best use all of the multiple estimates. We suggest to choose the set of factors that minimize the variance of (B.1), i.e., minimize

$$v(\hat{T}_{y,combined}) = \phi_{82}^2 v(\hat{T}_{y,82}) + \phi_{99}^2 v(\hat{T}_{y,99}) + \phi_{04}^2 v(\hat{T}_{y,04}) \quad (B.2)$$

If you take derivatives of (B.2) and solve for N_{82} , N_{99} and N_{04} , you find that their solution is proportional to the product of the variances for the other sample estimates, i.e., $\phi_k \propto \prod_{i \neq k} v(\hat{T}_{y,i})$

We can simplify this if we additionally assume that the variance for a given sample also inversely proportional to the sample size, i.e., $\phi_k \propto \prod_{i \neq k} n_i^{-1} (1 - f_i)$

Here f_i is the sampling fraction.

We account for the 1984 subsampling of the 1982 SPs who were unimpaired (SIC1 = 5) by assuming that N_k is proportional to a two-stage variance (Särndal 1992, eq 9.4.11), i.e.,

$$\phi_k \propto n_k^{-1} (1 - f_k) + \sum_H (n_{kh} / n_k)^2 n_{kh}^{-1} (1 - f_{kh})$$

where h is indexing the second stage strata. We calculate one N_k since both parts of the 1982 sample are needed to make a complete estimate.

The revised base weight is then the product of the original base weight and N_k .

2004 National Long Term Care Survey

Table C1: Calculating the Base Weights for the 95+

Age	Sample	Original BW	<i>n</i>	$(1 - f) / n$	<i>N</i>	Revised BW
95-99	1982-o	730.5909	23	0.005668	0.129611	94.6923
	1982-s	1,610.6208	160			208.7535
	2005-9	242.0218	1,180	0.000844	0.870389	210.6532
	total		1,363		1.000000	
100-104	1982-o	730.5909	10	0.030995	0.089287	65.2325
	1982-s	1,610.6208	25			143.8080
	1999-9	688.0678	104	0.009601	0.288233	198.3239
	2005-9	242.0218	224	0.004446	0.622480	150.6536
	Total		363		1.000000	
105+	1982-o	730.5909	4	0.052975	0.036059	26.3445
	1982-s	1,610.6208	16			58.0777
	1994-9	373.7185	100	0.009973	0.191538	71.5812
	1999-9	688.0678	121	0.008252	0.231477	159.2718
	2005-9	242.0218	282	0.003531	0.540926	130.9159
	Total		523		1.000000	

2004 National Long Term Care Survey

The following is an alternative method for assigning the base weight to units selected into the NLTCS sample in 1982 and 1984. This methodology was originally suggested by Census, but replaced with directions from CDS.

We can use the 1984 variable SCOMP84 (SAMP_CO on the NLTCS Utilities) to assign the base weight to sample people selected in 1982 and 1984. Table D1 associates values of SCOMP84 with the values of the base weight.

Table D1: Combined First and Second Phase Probabilities of Selection for 1984 SPs

SCOMP84	Description	$SI_{1&2,k}$
C	Completed a detail interview in 1982.	
C	Type C non-interview or Type A non-interview for the detailed interview.	730.5909
A	Type A non-interview for the screener and institutionalized before April 1, 1982.	[= 477.415818 H(101/66)]
B	Type C in 1982 and institutionalized before April 1, 1982.	
S	Not impaired in 1982.	1,610.6208 [= 730.5909 H(97/44)]
T	Aged-in for 1984.	730.5909

Note that SIC1 = 1, 2, 3, or 4 is equivalent to SCOMP84 = 'A', 'B' or 'C' and SIC1 = 5 is equivalent to SCOMP84 = 'S'. Jones (1984; p.16) and Waite (1988; p.1) explain the coding of SCOMP84.

In Table D1 'S' identifies the subsample of people taken in 1984.