

National Long-Term Care Survey: Assigning and Using the Pseudo Strata and Halfsample Codes

May 15, 2006

The information contained in this document is an extract from the following:
Stephen Ash, Demographic Statistical Methods Division, Census Bureau, "DRAFT: National Long-Term Care Survey: Assigning and Using the Pseudo Strata and Halfsample Codes," March 20, 2006.

I. Introduction

The purpose of this memorandum is to describe an assignment of the pseudo strata and halfsample codes for all sample people (SPs) in the 1982-2004 National Long Term Care Survey (NLTC). These codes can be used in estimating design variances using a replication variance estimator for **any wave** of NLTC.

The codes mask the identity of the PSU (Primary Sample Unit) within the data user file, so data users should not be able to narrow the identification of any individual to a small geographic area, i.e., PSU. In short, the codes give users enough information to estimate variances while keeping the identity of individual SPs confidential.

We provide separate instructions for certainty and noncertainty PSUs. Here we say a PSU is certainty if the probability of selection in the first phase is one and is non-certainty if it is less than one.

II. Assigning Pseudo Strata and Halfsample Codes

We now describe the assignment of the pseudo strata and halfsample codes that should be completed.

Non-Certainty PSUs

The assignment of the pseudo strata and halfsample code is defined by PSU and is specified by the Attachment A. This assignment is within region and has been used over the course of the survey to estimate the design variances. **Please Note:** *Attachment A does not appear in this document since it provides PSU information used in the conduct of the NLTC.*

Certainty PSUs

For the certainty PSUs, we will randomly assign the sample people of the 39 certainty PSU to one of the 41 new pseudo strata. Within pseudo strata, sample persons will also be assigned to one of the two halfsample codes. Identify certainty PSUs as having a first phase probability of selection equal to 1.0.

To complete this assignment, complete the following steps:

1. Assign two random numbers to each sample person. Call them RN1 and RN2. Generate both random numbers from a uniform distribution on the interval (0,1) and independently from each other.
2. Sort by
 - < PSU
 - < S_YEAR (Sample Year)
 - < State / county
 - < Date of Death (DOD)
 - < RN1
3. Assign the variable HALFSAMP the values 1 to 2 to each sample person, e.g., 1, 2, 1, 2, ...
4. Sort by
 - < HALFSAMP
 - < PSU
 - < S_YEAR
 - < State / county
 - < DOD
 - < RN2
5. Assign the variable PSEUDOSTRAT the values 1 to 41 to each sample person, e.g., 1, 2, ..., 40, 41, 1, 2, ...

Use the instructions in Attachment B to define the variable Sample Year (S_YEAR). Sorting by S_YEAR will insure that the number of SPs from each incoming sample is approximately uniform within each pseudo strata and halvesample combination. Sorting by DOD will help insure that the number of SPs alive and deceased within any given wave is also approximately uniform.

A Short Explanation of the Coding

We now explain why the memo codes sample people in certainty and non-certainty PSUs differently. The sampling error for NLTCs can be divided into two parts: the variance associated with the sampling of the PSUs and the variance associated with the sampling of people within PSUs. We sometimes refer to the two parts as the PSU variance and the within PSU variance.

For the non-certainty PSUs, we assign all the people from the same PSU to the same pseudo strata and half samples so that we can apply a Balanced Repeated Replication (BBR) variance estimator. BBR estimates both the PSU variance and within PSU variance (Wolter 1985; p. 117).

For the certainty PSUs, we do not need to estimate the PSU variance because there is none. We only need to estimate the within PSU variance. We do this by randomly assigning people from the same PSU to different pseudo strata and half samples. Each replicate sample then has a random half sample of people that is used to estimate the within PSU variance.

III. How to use the Pseudo Strata and Halfsample Codes to Estimate Design Variances

We now present two ways to use the pseudo strata and halfsample coding to produce estimated design variances.

Estimating variances with Balanced Repeated Replication

With BRR use variable PSEUDOSTRAT as strata and HALFSAMP as the primary sampling unit (PSU). We note that a Hadamard matrix of dimension 108 can be found at <http://www.research.att.com/~njas/hadamard/>.

Some good general references on BRR variance estimation include Wolter (1985; p. 110), Särndal et.al (1992; p. 430) and Cochran (1977; p. 320).

Estimating variances with WesVar™

We now explain how the codes can be used to estimate variances using Wesvar™ version 4.2. Enter the name of the variable containing PSEUDOSTRAT, into the field “VarStrat” of the “Create Weights” screen. Likewise enter the variable HALFSAMP as the “VarUnit”. Westvar™ requires that VarUnit be coded as 1 and 2, and VarStrat be coded as consecutive numbers starting from 1. The coding of this memorandum satisfies those requirements. Also select “BRR” for the Replication Method option. See Westat (2002) as a reference for Wesvar™.

References

- Ash, S. (2003). "1999 National Long-Term Care Survey: Instructions for Assigning and Using Pseudo Strata and Halfsample Codes," Census Bureau Memorandum from Stephen Ash for Documentation, dated December 9, 2003.
- Cochran, W. (1977). *Sampling Techniques*, John Wiley & Sons, New York.
- Särndal, C.E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*, Springer-Verlag, New York.
- Westat (2002). "Westvar 4.2 User's Guide,"
<http://www.westat.com/wesvar/about/WV4.2%20Manual.pdf>.
- Wolter, K.M. (1985). *Introduction to Variance Estimation*. Springer-Verlag, New York.

National Long Term Care Survey

Use Table B1 to define the variable Sample Year (S_YEAR) for all SPs in certainty PSUs. This variable identifies SPs who have the same base weight.

Table B1: Assignment of Sample Year

If then S_YEAR =
SCOMP84 = A, B, or C	1982-6
SCOMP84 = S	1982-S
SCOMP84 = T	1984-6
SIC2 = 6	1989-6
SIC3 = 6	1994-6
SIC3 = 9	1994-9
SIC4 = 6	1999-6
SIC4 = 9	1999-9
SIC5 = 6	2004-6
SIC5 = 9	2004-9

The variable SCOMP84 represents the 1984 sample components.