UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

Signed – April 14, 2005

MEMORANDUM FOR: THE RECORD

From: Stephen Ash
Mathematical Statistician
Demographic Statistical Methods Division

Subject: National Long-Term Care Survey: Calculation of Design Variances

The purpose of this memorandum is to describe recommended procedures for calculating design variances for the National Long Term Care Survey (NLTCS). By design variance we refer to the variability of the estimators with respect to randomness of the sampling from a finite universe. The variance estimator we describe can be used in conjunction with the sampling weights that are provided with the NLTCS data. The NLTCS is sponsored by the National Institute of Aging and conducted by the Duke University Center for Demographic Studies.

Before describing the variance estimator, we will first review some features of the NLTCS sample design that are important to estimating variances.

Sample Design

NLTCS is a two stage longitudinal survey where the first stage units are a county or a group of counties and the second stage units are people aged 65 years old or older. The original sample of Primary Sample Units (PSUs) has not changed since the inception of the survey in 1982. Within each PSU and for each 5 year wave, a sample of people aged 65 to 69 years old is selected to maintain the overall sample. A supplementary sample is also selected to maintain the sample of people aged 95 years old or older.

The PSUs are of two types – certainty and non-certainty. We say a PSU is certainty if the probability of selecting it is 1.0 and is non-certainty if the probability of selecting it less than 1.0. The NLTCS non-certainty PSUs were originally selected with probability proportional to size and one PSU per strata. Specifically NLTCS has 134 non-certainty PSUs and 39 certainty PSUs.

For estimating variances Ash (2005) has paired non-certainty PSUs into 67 pseudo strata. We therefore treat the paired PSUs as being selected two PSU per stratum. The sample persons (SPs) of the 39 certainty PSUs were randomly assigned to one of 41 pseudo strata and then further randomly assigned to one of two PSUs within the pseudo strata. The random assignment of SPs from certainty PSUs will permit a variance estimator to provide an estimate of the within PSU variance of those certainty PSUs. At the end there are 108 pseudo strata with two PSUs per each strata. The names for the two variables are P_STRATA which identifies the pseudo strata and HALFSAMP which identifies the PSU within each pseudo stratum.

Estimating variances with Balanced Repeated Replication

Balanced Repeated Replication (McCarthy 1966), (Fay 1984) is a method of creating many "half samples" which can be used to estimate the sampling variance of an estimator with sample design weights. The half samples are created by choosing one of the two PSUs of each pseudo stratum to represent the entire pseudo stratum. We say the half samples are "balanced" because we use the rows of a Hadamard matrix to define each half sample. A Hadamard matrix has values of '+' or '-' and is balanced because all of the rows are orthogonal to each other.

To form the half samples, let one of the two PSUs in each pseudo stratum be the '+' PSU and let the other be '-' PSU. Then associate the rows (r) of the Hadamard matrix with each of the half samples and associate the columns (c) with the pseudo strata. Include the '+' PSU from the c^{th} pseudo strata in the r^{th} half sample if the corresponding value from the Hadamard matrix is '+', and include the '-' PSU otherwise. At the end of this procedure there will be the same number of half samples as there are pseudo strata since the Hadamard matrix has the same number of rows and columns. We note that a Hadamard matrix of dimension 108 can be found at www.research.att.com/~njas/hadamard/.

Following Särndal et.al (1992; p. 430), estimate the statistic of interest for each half sample as you would with the entire sample, except multiplying each SP with a factor of two. We denote this half sample estimate as $\hat{t}_a$, where *a* is indexing the half samples. The estimate of the variance then is

$$\hat{v}_{BH} = \sum_{a=1}^{A} \left(\hat{t}_a - \hat{t}_{BH} \right)^2$$

where $\hat{t}_{BH} = \frac{1}{A} \sum_{a=1}^{A} \hat{t}_a$ and *A* is the number of half samples.

Note that one by-product of the balancing is that $\hat{t}_{BH}$ equals the estimate using the entire sample for totals and most other statistics.

Some good general references on BRR variance estimation include Wolter (1984; p. 110-145) and Särndal et.al (1992; p. 430-437).

In the instructions for calculating the sample weights of each wave (Jones 1982, 1984), (Waite 1989, 1994), (U.S. Census 2005) a second stage ratio adjustment factor should have been applied as the last weighting adjustment factor for each of the different weights. With a replicate variance estimator this second stage ratio adjustment is also applied to each replicate. Although the known totals are provided in the weighting specifications, we recommend not recalculating the second stage ratio adjustment factor for each replicate with estimates prior to and including 1999.

Estimating variances with WesVar®

We now explain how to use Wesvar® version 4.2 to calculate the BRR variances. Enter the name of the variable containing P_STRAT, 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 P_STRAT and HALFSAMP satisfy those requirements. Also select "BRR" for the Replication Method option. See Westat (2002) as a reference for Wesvar®.

References

Ash, S. (2005). "National Long-Term Care Survey: Instructions for Assigning and Using Psuedo Strata and Halfsample Codes," Census Bureau Memorandum for Documentation from Stephen Ash, dated Draft.

Fay, R.E. (1984). "Some Properties of Estimates of Variance Based on Replication Methods," Proceedings of the American Statistical Association, Section on Survey Methods, 495-500.

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.

McCarthy, P.J. (1966). "Replication: An Approach to the Analysis of Data from Complex Surveys," Vital Health Statistics, 2, 14, National Center for Health Statistics. Washington D.C.: U.S. Government Printing Office.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). Model Assisted Survey Sampling, Springer-Verlag, New York.

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. (1994). "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.

U.S. Census (2005) "1999 National Long-Term Care Survey: Weighting for Cross-sectional Estimation," draft 2005.

WesStat (2002). “Westvar® 4.2 User’s Guide,” (www.westat.com/wesvar/).

Wolter, K.M. (1982). Introduction to Variance Estimation, Spring-Verlag, New York, NY.

cc:

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

S:\LTC\LTC2004\memos\nltcs2004_variance_instructions_final.wpd