Calibration Weights for Estimators of Longitudinal Data with an Application to the National Long Term Care Survey

Stephen Ash
Bureau of the Census, 4700 Silver Hill Road, Rm 3727-3, Washington D.C. 20233-8700

Abstract

We consider a set of weights appropriate for estimating differences of a given variable of interest between two separate waves of the National Long-Term Care Survey (NLTCS). The NLTCS is a nationally representative longitudinal survey of people aged 65 years old and older. The initial wave of the survey was in 1982 and it has been conducted every five years since 1984. The survey examines trends in disability, impairment, and mortality.

In the cross-sectional weighting of many surveys we use a known set of totals to create second stage ratio adjustment factors, i.e., to post-stratify. This factor helps to correct the estimates for coverage problems and reduces the sampling variance of the resultant estimators.

With an estimator of the difference of some value at two different times, it is not obvious how you would use the two sets of known totals that are available. For example, with NLTCS we have known totals for age/sex/race for both 1994 and 1999. Instead of only using one set of known totals, we consider the problem as a calibration problem which leads to a solution for two sets of known totals. This solution adjusts individual sample units who report at both times for both sets of known totals.

Keywords: Longitudinal survey, non-interview

1. Introduction

The motivation of this paper is the development of estimators that use longitudinal data from the National Long Term Care Survey (NLTCS). In this development we will show how to use auxiliary data that are available for the two time points of interest to improve estimators. Here we define auxiliary information as those values of variables known for all units in the universe.

We will discuss the "usual" methods for using auxiliary data with longitudinal estimators and will also discuss two alternative calibration methods. In conjunction with the calibration methods we will also discuss how to account for non-interviews. Lastly we will provide some results that apply the discussed methods to data from the 1994 and 1999 NLTCS.

1.1 Sample Design of NLTCS

The NLTCS 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 NLTCS is sponsored by a grant from the National Institute of Aging (5 R01 AG007198-15 for the 1999 survey and 5 U01 AG007198-18 for the 2004 survey) and conducted by the Duke University Center for Demographic Studies (CDS).

The universe for NLTCS is the civilian population of persons 65 or older in the U.S. The sampling frame is an extract from the Enrollment Database, which is an administrative record of the Center for Medicare and Medicaid Services (CMS) that is used to maintain data on payment status, demographic characteristics, and other information necessary to administer the Medicare program.

Because the main focus of the survey is on measuring disability, the survey uses a screener interview to identify sample persons that are disabled or institutionalized. Sample persons who are disabled or institutionalized then receive a detailed interview. A subsample of "healthy" persons is also selected and receive the detailed interview so that all eligible persons have a non-zero probability of being selected for the detailed interview.

Although NLTCS is a longitudinal survey, not every sample person is interviewed in every wave. We only include a sample of the people who were not disabled and not institutionalized in a previous wave in the next wave.

1.2 Statistics for Longitudinal Data

We now define two statistics used to compare values from different waves. We begin with an example of each from NLTCS.

Q1: How many more people aged 65 years old and older were disabled in 1999 than in 1994?
Q2: How many people that were disabled in 1994 were also disabled in 1999?

We answer each of these questions by applying different statistics to the longitudinal data.

We refer to the statistic related to Q1 as a difference and define it as

$$D = \sum_{U_{t=2}} y_{2k} - \sum_{U_{t=1}} y_{1k} \ .$$

Here we define $U_{t=1}$ and $U_{t=2}$ as the universe at times t=1 and t=2, respectively. As in AAPOR (2004) we say a unit is eligible if it is in the universe of interest. The variables y_{1k} and y_{2k} denote the variables of interest where y_{1k} is an observation at t=1 and y_{2k} is an observation at t=2. The subscript k indexes the eligible units.

We can also divide $U_{t=1}$ and $U_{t=2}$ into two parts. The first part is the intersection of $U_{t=1}$ and $U_{t=2}$, i.e., the units that are eligible at both t=1 and t=2, which we will refer to as E. The second part is the units that are eligible for only one of the times, i.e., $M_{t=1}$ and $M_{t=2}$. For NLTCS $M_{t=1}$ represents those persons eligible at t=1 and deceased at t=2 and $M_{t=2}$ represents those persons not yet 65 at t=1 and then at least 65 at t=2.

With this notation we can express D as

$$D = \sum_{E} \left(y_{2k} - y_{1k}\right) + \left(\sum_{M_{t=2}} y_{2k} - \sum_{M_{t=1}} y_{1k}\right) \quad .$$

The statistic of Q2 we will refer to as a longitudinal intersection because it can represent the intersection of two domains. We define the spell statistic or our term, longitudinal intersection, as

$$L = \sum_{E} y_{1k} y_{2k} \quad .$$

In Q2 we code both variables of interest y_{1k} and y_{2k} equal to 1 when the eligible unit is disabled and 0 when non-disabled.

Note that the intersection statistic can only be estimated from a longitudinal survey. We need the observations of y_{1k} and y_{2k} from the same unit. We also do not allow recall, i.e., the collection of y_{1k} at t=2, because of all the problems associated with retrospective reporting.

One estimator of D is

$$\begin{aligned}\hat{D}_1 &= \sum_{s_{t=2}} w_{2k} y_{2k} - \sum_{s_{t=1}} w_{1k} y_{1k} \\ &= \left(\sum_{r_2} w_{2k} y_{2k} - \sum_{r_1} w_{1k} y_{1k}\right) \\ &\quad + \left(\sum_{m_2} w_{2k} y_{2k} - \sum_{m_1} w_{1k} y_{1k}\right)\end{aligned}$$

In our notation $s_{t=1}$ and $s_{t=2}$ denote the sample at t=1 or t=2, respectively. Also r_1 denotes the units eligible at t=1 and t=2 and respond at t=1. We similarly define r_2. Note that units who are eligible and respond at both t=1 and t=2 are in both r_1 and r_2. Likewise m_1 and m_2 denote the respondents who were only eligible at t=1 or t=2, respectively. The variables w_{1k} and w_{2k} are weights specific to t=1 and t=2. We will say more about the weights in the next section.

Working from our alternate expression for D we see that we can estimate D by only using those eligible units that respond at both times in part of the estimator relating to the units in E, i.e.,

$$\hat{D}_2 = \sum_{r} w_k \left(y_{2k} - y_{1k}\right) + \left(\sum_{m_2} w_k y_{2k} - \sum_{m_1} w_k y_{1k}\right)$$

In $\hat{D}_1$ we use all of the sample units that are eligible at both times and respondents at least once (r_1 or r_2) to estimate the difference for the units of E. In $\hat{D}_2$ we only use those who are eligible and respond at both times ($r = r_1 \cap r_2$). We say that the weights w_{1k} and w_{2k} of $\hat{D}_1$ are cross-sectional and the weight w_k of $\hat{D}_2$ is longitudinal.

An estimator of L is

$$\hat{L} = \sum_{r} w_k y_{1k} y_{2k}$$

where the weight w_k must be a longitudinal weight as in $\hat{D}_2$ and we only use those units who are both eligible and who respond at both times.

Note that Kalton and Citro (1994) refer to the general analysis of D as a "measurement of gross change" and of L as "relationship between variables across time."

2. NLTCS Sample Weighting

2.1 Overview of the Weighting

The weighting of NLTCS accounts for several random processes which include:

- The first stage sample of PSUs (counties or groups of counties) originally selected in 1982.
- The second stage sample of Medicare enrollees. This includes the original 1982 sample of persons aged 65 years old and older. It also includes the samples we have selected in 1984 and every 5 years since. The 5 year samples update the sample by selecting persons who have turned 65 years old since the most previous sample.
- Subsampling of longitudinal sample persons who were previously healthy.
- Screener and detailed non-interviews.

In our examples later, we will focus on estimates about disability, so the weighting we describe is approximately how the CDS Screener Weight was calculated. This weight is available on the NLTCS public use dataset and the NLTCS Utlilties (2001). The CDS Screener Weight adjusts all sample units that complete a screener interview for screener non-interviews, uses people who screen-out (were not disabled or institutionalized and therefore did not receive a detailed interview) as completed interviews and then for sample units who screen-in, adjusts detailed interviews for detailed non-interviews.

We note that the weights we calculated are different from those available on the NLTCS Utilities. We were privy to more up-to-date information than was available in 1994 and 1999 and we also worked to keep some of the computation within the research simple.

2.2 Non-Interview Adjustments

The non-interview adjustments for w_{1k} and w_{2k} are much like non-interview adjustments for other cross-sectional weights. Usually a non-interview factor is applied to the sampling interval to account for non-interviews where the non-interview factor represents the inverse of the probability of being a completed interview for the given eligible sample unit. There are two popular methods for calculating the non-interview probability. The commonly used method is to estimate the value from the sample within non-interview cells. The second is to use a model to estimate the value from the sample (Folsom and Witt 1994), (Rizzo et.al 1994).

The non-interview adjustment for the longitudinal weights we suggest are derived as in the review of Survey of Income and Program Participation non-interview adjustments given by Rizzo et.al. (1994; p. 422) and Chapman et.al. (1986). The non-interview adjustment for the units of r can be summarized as follows. Given we only use units that are eligible and respond to both waves to represent E, we apply two adjustments to the units of r. The first adjusts for the t=1 non-interviews and the second for the t=2 non-interviews given the t=1 non-interviews. For sample units of m_1 and m_2, we applied a single non-interview adjustment.

In our application of NLTCS we applied separate non-interview adjustments for the screener and detailed interviews at both t=1 and t=2. Therefore we applied four different non-interview adjustments to the units of r.

In this paper we calculated the non-interview adjustments separately for the groups $M_{t=1}$, $M_{t=2}$ and E. Within each group we calculated the non-interview adjustment within non-interview cells. Both the cross-sectional and longitudinal non-interview adjustments used the same non-interview cells. For the screener and detail non-interview adjustment we defined cells as all possible combinations of the following categorical variables: region (4 values), urban/rural, and age (5-year groups).

2.3 Calibration

The calibration for cross-sectional weights used with demographic surveys is often referred to as poststratification (Bethlehem 1988) or a ratio estimator adjustment. Here we apply a factor to the weight that is the ratio of the known total and its estimated value.

Within the usual calibration framework (Deville and Särndal 1992) we can equivalently and more generally describe poststratification as follows. Say that x_{1k} and x_{2k} are known and available auxiliary data for t=1 and t=2, respectively. We use this auxiliary information by finding new cross-sectional weights w_{1k} and w_{2k} based on the original design weight d_k, the sampling interval, and the subsampling and non-interview adjustments c_{1k} and c_{2k} that are specific to each wave. Specifically we begin by finding weights w_{1k} for t=1 that minimize ϕ_{1a} then we find weights w_{2k} for t=2 that minimize ϕ_{1b}. Tables 2 and 3 provide the exact calibration equations used throughout the paper.

Table 2: Calibration Equations for Cross-sectional Weights

Method	Calibration Equation
Usual two step method	$\phi_{1a} = \frac{1}{2}\sum_{r_1,m_1} \frac{(w_k - c_{1k}d_k)^2}{c_{1k}d_k} - \left(\sum_{r_1,m_1} w_{1k}\mathbf{x}_{1k} - \sum_{U_{t=1}} \mathbf{x}_{1k}\right)'\lambda$
	$\phi_{1b} = \frac{1}{2}\sum_{r_2,m_2} \frac{(w_k - c_{2k}d_k)^2}{c_{2k}d_k} - \left(\sum_{r_2,m_2} w_{2k}\mathbf{x}_{2k} - \sum_{U_{t=2}} \mathbf{x}_{2k}\right)'\lambda$
Alternative one step method	$\phi_2 = \frac{1}{2}\sum_{r_1,r_2,m_1,m_2} \frac{(w_k - d_k)^2}{d_k} - \left(\sum_{r_1,m_1} c_{1k}w_k\mathbf{x}_{1k} - \sum_{U_{t=1}} \mathbf{x}_{1k}\right)'\lambda - \left(\sum_{r_2,m_2} c_{2k}w_k\mathbf{x}_{2k} - \sum_{U_{t=2}} \mathbf{x}_{2k}\right)'\gamma$

Table 3: Calibration Equations for Longitudinal Weights

Method	Calibration Equation
Usual two step method	$\phi_{3a} = \frac{1}{2}\sum_{r,m_1}\frac{\left(w'_k - c_{1k}d_k\right)^2}{c_{1k}d_k} - \left(\sum_{r,m_1} w'_k\mathbf{x}_{1k} - \sum_{U_{t=1}}\mathbf{x}_{1k}\right)'\lambda$ $\phi_{3b} = \frac{1}{2}\sum_{r,m_2}\frac{\left(w_k - c_{2k}w'_k\right)^2}{c_{2k}w'_k} - \left(\sum_{r,m_2} w_k\mathbf{x}_{2k} - \sum_{U_{t=2}}\mathbf{x}_{2k}\right)'\gamma$
Alternative one step method	$\phi_4 = \frac{1}{2}\sum_{r,m_1,m_2}\frac{\left(w_k - c_{1k}c_{2k}d_k\right)^2}{c_{1k}c_{2k}d_k} - \left(\sum_{r,m_1} w_k\mathbf{x}_{1k} - \sum_{U_{t=1}}\mathbf{x}_{1k}\right)'\lambda - \left(\sum_{r,m_2} w_k\mathbf{x}_{2k} - \sum_{U_{t=2}}\mathbf{x}_{2k}\right)'\gamma$

In this paper we also consider the following alternative calibrated estimator where the weights w_k are those that minimize ϕ_2. One of the goals of ϕ_2 is to retain the subsampling factors and non-interview adjustments specific to t=1 and t=2. For example, if a unit had a subsampling factor of 2 in 1994 and none in 1999, we wanted the final weight for 1994 to be twice as large as the weight for 1999.

To calibrate the longitudinal weight w_k it is customary to apply two adjustments, each after the two non-interview adjustments. We first calculate the non-interview adjustment c_{1k} that accounts for non-interviews at t=1. Next we find the intermediate weight w'_k by minimizing ϕ_{3a} and using $c_{1k}\ d_k$ as the survey weight. We then calculate a second non-interview adjustment which accounts of non-interviews at t=2 given the non-interviews of t=1. Finally, we calculate the final weight by minimizing ϕ_{3b} to find the w_k that is closest to $c_{2k}\ w'_k$.

In this paper we also consider the following alternative calibrated weight. We calculate this weight w_k by minimizing ϕ_4. This weight calculates both non-interview adjustments prior to a single calibration step. An advantage of this weight over the weights generated by the minimizing of ϕ_{3a} and ϕ_{3b} is that the estimated totals x_{1k} using w_k will equal the known totals. This is not the case with the two step calibration since the second calibration of ϕ_{3b} changes the weights for units in E after the calibration of ϕ_{3a}.

Specific expressions for the weights resulting from calibration equations ϕ_2 and ϕ_4 are included in Ash (2005). Within this paper we also used the squared error distance function or case 1 in Deville and Särndal (1992). Although other distance functions can guarantee a positive weight, the squared error distance is often easier to minimize and solve with a closed solution.

In the estimates later we used auxiliary data from both 1994 and 1999 in the form of population totals. The totals were for the crossed categories of age (5-year categories), race (black/non-black), sex (male/female).

3. Variance Estimation

Since NLTCS has a multi-stage sample design where a sample of Primary Sampling Units (PSUs) is selected in the first stage, we used Balanced Repeated Replication (McCarthy 1966), (Fay 1984) method to estimate variances and re-calculated the non-interview adjustments and calibration for each replicate.

4. Results

Table 4 provides the estimates of D for several variables of the 1994 and 1999 NLTCS. The value in parentheses after the estimate is the estimated standard error. We derived all of the estimates from the single sample of the 1994 and 1999 NLTCS, therefore our results cannot be used to discuss the bias or mean squared error of these estimators.

The disability variables we used are key in NLTCS. A person is either non-disabled, disabled or institutionalized. A person is categorized by the NLTCS interview as disabled if they received help with one or more Activities of Daily Living (ADL) or Instrumental Activities of Daily Living (IADL). We present a handful of the possible estimates of interest. We now list some observations about the estimates of D.

Note 1: All of the types of calibration reduce the estimated variance as compared with not applying any calibration.

Note 2: Calibrating the cross-sectional weights for $\hat{D}_1$ with ϕ_{1a} and ϕ_{1b} produced estimates with smaller estimated variances than calibrating with ϕ_2. We realized after producing this result that the goal of the alternative calibration ϕ_2 – maintaining the non-interview and subsampling factors c_{1k} and c_{2k} – preserved those sources of randomness in the final weight.

Note 3: Calibrating the longitudinal weights for $\hat{D}_2$ with ϕ_4 produced estimates with smaller estimated variances than calibrating with ϕ_{3a} and ϕ_{3b}. The calibration of ϕ_4 was also simpler because it required one calibration step instead of two.

Note 4: Although the estimator $\hat{D}_1$ used more observations than $\hat{D}_2$, it did not have a smaller estimated variance, as we expected. We started wit the assumption that the variance of $\hat{D}_2$ would be smaller because it used more observations. However we think the variance of $\hat{D}_1$ was actually smaller because the weights were simpler – all sample units of E had the same weight, so units with no change contributed nothing to the estimated difference and also nothing to the variance. On the other hand with the cross-sectional weights, units with no change contributed the difference between w_{1k} and w_{2k} and thereby also contributed a positive value to the estimated variance.

Note 5: The results for the variance estimates appear exactly the same the estimated standard errors were smaller when calibrating with ϕ_4 than with ϕ_{3a} and ϕ_{3b} after the thousandths digit.

Note 6: We found it simpler to implement the weights associated with ϕ_2 and ϕ_4 than those having simpler calibration adjustments that are applied in two steps. Applying one complicated calibration adjustment is easier than two because each step requires the same summing for the estimated totals and then merging those totals back to the observations. This summing and merging is done once with the weights associated with ϕ_2 and ϕ_4 and twice with the others.

5. Conclusions

We have shown how to use auxiliary data available for both t=1 and t=2 in conjunction with longitudinal sample data. We have also seen how the handling of the real world problem of non-interviews and how we adjust the sample weights for them impacts how we use the auxiliary data when calibrating.

Of the methods considered in our paper the best method for estimating D and L is calibrating the longitudinal weights with ϕ_4. The estimates had approximately the same or smaller estimated variance when compared with the other estimators. We also liked it because it was simple to apply.

We were disappointed that the estimator $\hat{D}_2$ did not perform better than $\hat{D}_1$. We understand that units that are observed twice contribute less to the variance, however it still seems counter-intuitive that an estimator that uses less of the observed sample has a smaller variance than one that uses more.

This paper reports the results of research and analysis undertaken by the U.S. Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

References

American Association for Public Opinion Research (2004). "Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys", Lenexa, Kansas: AAPOR

Ash, S. (2005). Census Bureau Memorandum "2004 National Long Term Care Survey: Estimators for Longitudinal Sample Designs," from Stephen Ash to Documentation dated draft.

Bethlehem, J.G. (1988). "Reduction of Nonresponse Bias Through Regression Estimation," Journal of Official Statistics, 4, 251-260.

Chapman, D.W., Bailey, L., and Kasprzyk, D. (1986). "Nonresponse Adjustment Procedures at the U.S. Bureau of the Census," Survey Methodology, 12, 161-181.

Deville, J-C. and Särndal, C.E. (1992). "Calibration Estimators in Survey Sampling," Journal of the American Statistical Association, 87, 376-382.

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

Folsom, R. and Witt, M.B. (1994). "Testing a New Attrition Nonresponse Adjustment Method for SIPP," Proceedings of the American Statistical Association, Section on Survey Methods, 428-433.

Kalton, G. and Citro, C.F. (1993). "Panel Surveys: Adding the Fourth Dimension," Survey Methodology, 19, p. 205-219.

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.

McCarthy, P.J. (1966). "Replication: An Approach to the Analysis of Data from Complex Surveys," Vital Health Statistics, ser. 2, no. 14, National Center for

Health Statistics. Washington D.C.: U.S. Government Printing Office.

National Long Term Care Survey Utilities (2001). Data, Documentation, and Extraction Utilities, copyright 2001-2005 Unicon Research Corporation, Santa Monica, CA, with funding from the National Institute of Aging under grant 5 R44 AG19117-03.

Rizzo, L., Kalton, G., Brick, M., and Petroni, R. (1994). "Adjusting for Panel Nonresponse in the Survey of Income and Program Participation," Proceedings of the American Statistical Association, Section on Survey Methods, 422-427.

Table 4: Estimates of D - Difference in Totals between 1999 and 1994 (in thousands)

	$\hat{D}_1$ w/ no calibration	$\hat{D}_1$ calibrated w/ ϕ_{1a} and ϕ_{1b}	$\hat{D}_1$ calibrated w/ ϕ_2	$\hat{D}_2$ calibrated w/ ϕ_{3b} and ϕ_{3b}	$\hat{D}_2$ calibrated w/ ϕ_4
Total	2,016 (212)	2,124 (0)	2,124 (0)	2,700 (123)	2,124 (0)
Non-disabled	1,941 (240)	2,152 (149)	2,357 (152)	1,850 (174)	1,661 (151)
Disabled (IADL or ADL)	273 (132)	177 (131)	52 (135)	888 (126)	681 (127)
IADL Only	-375 (81)	-356 (79)	-377 (78)	-126 (76)	-164 (75)
5-6 ADLs	189 (62)	151 (59)	108 (59)	200 (55)	145 (56)

Table 5: Estimates of L (in thousands)

	$\hat{L}$ calibrated w/ ϕ_{3a} and ϕ_{3b}	$\hat{L}$ calibrated w/ ϕ_4
Disabled in 1994 and disabled in 1999.	18,777 (118)	18,778 (118)
Non-disabled in 1994 and disabled in 1999.	9,342 (66)	9,344 (66)
Non-institutionalized in 1994 and institutionalized in 1999.	1,234 (59)	1,234 (59)
Non-disabled in 1994 and institutionalized in 1999.	775 (54)	775 (54)