COMBINING SAMPLES VS. CUMULATING CASES: A COMPARISON OF TWO WEIGHTING STRATEGIES IN NLSY97

Colm O'Muircheartaigh, Steven Pedlow, NORC
Colm O'Muircheartaigh, National Opinion Research Center, 55 E. Monroe, Chicago, IL 60603

Key Words: NLSY97, weights, combining samples, cumulating cases.

1. Introduction. The National Longitudinal Survey of Youth (NLSY97) is the latest in a series of surveys sponsored by the U.S. Department of Labor (DoL) to examine issues surrounding youth entry into the work force and subsequent transitions in and out of the work force. The NLSY97 is following a cohort of approximately 9,000 youths who completed an interview in 1997 (the base year). These youths were between 12 and 16 years of age as of December 31, 1996, and are being interviewed annually using a mix of some core questions asked annually and varying subject modules. We will compare two different weighting strategies for the first three rounds of NLSY97 data.

In order to improve the precision of estimates for minority youths, the overall study design for NLSY97 included a large oversample of Hispanic youths and non-Hispanic black youths. The overall design resulted in one large screening sample of over 90,000 housing units to generate youth participants for NLSY97. These housing units were drawn from two independent area-probability samples: 1. a cross-sectional (CX) sample designed to represent the various segments of the eligible population in their proper population proportions, and 2. a supplemental (SU) sample designed to produce, in the most statistically efficient way, the required oversamples of Hispanic youths and non-Hispanic black youths. This paper's main concern is with the construction of sampling weights for estimating population characteristics using both samples together. The paper gives more detailed descriptions of the two samples in Section 2. Section 3 describes the weighting methodology used not only in the first two rounds of NLSY97, but also in the earlier NLS79 cohort; the approach was based on combining weights constructed separately for each sample. Section 4 describes an alternative methodology that cumulates the selection probabilities for cases across the two samples. Section 5 compares the two approaches theoretically, Section 6 compares the two approaches empirically, and Section 7 presents our conclusions.

2. The Two Samples. The CX and SU samples were both selected by standard area-probability sampling methods. Sampling was in three stages: primary sampling units (PSUs), consisting mainly of Census metropolitan statistical areas (MSAs) or single counties; secondary sampling units that were segments consisting of single census blocks or clusters of neighboring blocks; and housing units (HUs) as tertiary sampling units. All eligible youths in each household were then selected for interviewing and testing.

However, the two samples did differ in how probabilities were assigned for this multi-stage area-probability sampling. The CX sample was an approximately equal-probability sample of households. For each of the first two stages (PSU and segment), a weighted systematic sample was selected in which the weight was equal to the number of housing units. The third stage (HUs) selected HUs within selected segments at differing rates in order to equalize the housing unit selection probabilities. Thus, each housing unit selected had approximately the same selection probability.

The SU sample was designed to oversample Hispanic and non-Hispanic black youths. Again, for each of the first two stages (PSU and segment), a weighted systematic sample was selected, but this time, the weight used was a weighted sum of 1990 Decennial Census counts of Hispanic youths and non-Hispanic black youths. At the housing unit selection stage, the segments were divided into two strata: a high minority youth stratum, and a low minority youth stratum. Within each stratum, HUs were again selected at rates that attempted to equalize the housing unit selection probabilities. However, the equalized selection probability in the high minority youth stratum was ten times the equalized selection probability in the low minority youth stratum.

Table 1 shows the sample sizes for each round by sample and race/ethnicity.

3. Combining the CX and SU Samples. The NLSY program has been using the same approach to weighting since the 1979 cohort was established. This method, used for NLSY97 rounds 1-3 (as well as NLS79), is based on calculating housing unit sampling weights separately within each of the two samples (CX and SU). These weights treat the CX and SU screening samples as stand-alone samples. Within each sample, the weights for a given domain (such as Hispanic males) were designed to sum to the population size for that group. In order to permit analysis of both samples together, these sample-specific weights were adjusted. To maintain the characteristic that the weights from both samples together sum to the population size (rather than each sample independently), the CX weights were multiplied by λ ($0<\lambda<1$), and the SU weights were multiplied by 1-λ in producing estimators based on both samples together:

$$\hat{\theta} = \lambda\hat{\theta}_c + (1-\lambda)\hat{\theta}_s$$

in which $\hat{\theta}_c$ represents a statistic derived from the CX sample and $\hat{\theta}_s$ represents the corresponding statistic from the SU sample. Because the two samples are independent, the optimum λ for a weight of this form is proportional to the relative effective sample size in the CX sample:

$$\lambda = \frac{n_c / d_c}{n_c / d_c + n_s / d_s}$$

$$1-\lambda = \frac{n_s / d_s}{n_c / d_c + n_s / d_s}$$

in which n_c and n_s are the nominal sample sizes for the CX and SU samples and d_c and d_s represent the design effects for the estimators from each sample. It is inconvenient to use the design effects themselves, since they will vary from one variable to the next. Instead, a general factor was used (one plus the squared coefficient of variation of the weights within each sample), as was done for NLS79; this factor captures the impact of unequal weighting on the sample efficiency:

$$\hat{d}_c = 1 + \left[CV(W_i \in \text{ CX})\right]^2$$

$$\hat{d}_s = 1 + \left[CV(W_i \in \text{ SU})\right]^2$$

The calculation of λ was carried out separately for each race/ethnicity by sex combination (e.g., Hispanic males). Race/ethnicity was defined as Hispanic, non-Hispanic black, or non-Hispanic, non-black. At the level of individual members of the sample, the weight after combining samples (W*) is the sample member's weight from the previous step (W) times the relevant combination factor ($\hat{\lambda}_\beta$), where β signifies the sex by race/ethnicity domain:

$$W_i^* = \hat{\lambda}_\beta W_i \qquad , \; i \in \text{ CX}$$

$$= (1-\hat{\lambda}_\beta)W_i, \; i \in \text{ SU}$$

This method is a sample-based weighting system, in that for all the individuals in a particular race/ethnicity by sex combination, it is the sample to which the case belongs (CX or SU) rather than the characteristics of the individual case that determines the modification of the weights. No knowledge of the sample design (other than the weights) is used in determining the combination. One important feature is that, within each sample and domain, the relative relationship of case weights remains the same; i.e. if a case has a larger weight before the samples are combined, the case will still have a larger weight after combining the samples. This "combining samples" methodology can also be used to combine samples from different surveys altogether.

4. Cumulating Cases Across the Two Samples. As part of the analysis of NLSY97 Round 2, we also computed an alternative set of weights. The chief alternative to the "combining samples" approach described in section 3 would be to calculate each case's overall probability of selection (into either the CX or SU screening samples), and then to use weights inversely proportional to these overall probabilities of selection. These are Horvitz-Thompson weights.

As Kish observed in connection with the 1979 study (see Appendix 2 in Frankel et al., 1983), the two approaches necessarily produce converging results, provided that the required selection probabilities can be reconstructed for both samples. However, calculating the CX probability of selection for the cases selected into the SU sample (and the SU selection probability for the cases selected into the CX sample) is not necessarily simple. Moreover, as Kish argued then, the Horvitz-Thompson approach may become more complicated when nonresponse adjustment factors are applied within each sample.

In this Horvitz-Thompson approach, the weights are determined across samples depending only on the overall selection probability (into either sample) of the individual element, giving a single unified set of weights for the cumulated cases. This approach is straightforward. The probability for a case to be in either sample is simply the sum of the probabilities to be in each sample because the samples are independently drawn. Thus, the base weight for a case is the inverse of the sum of sample selection probabilities for a case:

$$W_i^* = \frac{1}{P_i^{CX} + P_i^{SU}}$$

$$\begin{aligned} &where \\ &P_i^{CX} = sel.\ prob.\ for\ case\ i\ in\ CX\ sample, and \\ &P_i^{SU} = sel.\ prob.\ for\ case\ i\ in\ SU\ sample \end{aligned}$$

This is a case-based weighting system, and would not be easily used for separate samples because then the selection probabilities for the other sample might be very difficult to obtain. The weight for each case is based not on the sample from which it came but on the probability that it *could* have been selected into *either sample*. That is of course the probability that the case had of appearing in the data set.

5. Comparing the Two Strategies - Theoretical. Both of the strategies are unbiased, but they have potentially different impacts on variance. We first consider the implications of the two weighting systems on variances using the general impact of variation in weights on the variance of estimators. In Section 6, we will compare the outcomes using the two strategies on substantive estimators from NLSY97 data.

A well-known property of weights (Kish, 1965) is that arbitrary weights increase the variance by a factor 1+L where:

$$L = \frac{Var(W_i)}{\overline{W}^2} = [CV(W_i)]^2$$

It should be noted that $\hat{d}_C$ and $\hat{d}_S$ (estimated design effects) from section 3 above are both equivalent to L. We can then compare the two weighting strategies by comparing the effective sample sizes (n_{eff}) under each strategy, where:

$$n_{eff} = \frac{n}{1 + L}$$

The above effective sample sizes only account for the variability in the weights, but the other main factors in reducing effective sample sizes (sample design and clustering issues) will impact both weighting strategies equally if they are both calculated from the same round of data.

Tables 2-4 below compare the two strategies using Round 2 data. Each table has three columns. The first column shows the effective sample size when only the CX data are used; in this case, the SU interviews are not used. The second column shows the estimated gains from using the SU interviews with weights calculated on the basis of the combining samples strategy described in section 3; the third column shows the estimated gains using the cumulating cases strategy described in section 4.

Table 2 shows the comparison for the overall population. When the SU interviews are used with either weighting strategy, the coefficient of variation more than doubles. This indicates that the variability in the SU weights (under either strategy) is much greater than for the CX weights. The extra variability in the SU sample weights is due to the differential probabilities of selection at the third stage of selection – the selection of HUs. The added variability in the SU weights means that the 2110 SU interviews add only 695 to the effective sample size using the combining samples strategy. The cumulating cases strategy does better, adding 1012 to the effective sample size by using the 2110 SU interviews. The total effective sample size is 5% more than for the combining samples strategy, and gets 46% more of a gain from using the SU interviews.

Even larger gains arise when we use the *cumulating cases* weight for analyses of the minority youths, as shown in Tables 3 and 4. Table 3 shows that including the 1183 SU non-Hispanic black interviews adds 484 to the effective sample size of non-Hispanic black youths when the *combining samples* weights are used; the *cumulating cases* strategy adds 877 effective cases – an additional 393 above and beyond the 484. This increases the effective sample size for non-Hispanic black youths by 27% over the combining samples strategy.

Table 4 shows that the *combining samples* strategy adds 281 to the effective sample size of Hispanic youths when the 924 SU Hispanic interviews are included, while the *cumulating cases* strategy adds 657 from these same 924 interviews. Not only does this add 34% to the total effective sample size beyond the combining samples strategy, but it more than doubles the gain realized from the SU Hispanic interviews.

Tables 2-4 clearly show that the cumulating cases strategy has the larger effective sample size because the weights have a smaller coefficient of variation. The CX selection probabilities are almost all identical because the CX sample is designed to represent all segments of the eligible population in their proper population proportions. The combined probabilities are the sum of the CX and SU probabilities. By adding a constant to the SU selection probabilities, the range remains the same, but the variability of the inverse is decreased because the smallest probabilities are farther from zero. Of course, adding the more variable SU selection probabilities to the CX selection probabilities will increase the variability among the selection probabilities and base weights for the CX cases, but our results show that this countereffect for CX cases is smaller than the effect for SU cases. If the two samples had weights with more similar variabilities, the gain from the cumulating cases method would in general be less.

6. Comparing the Two Strategies – Empirical. The theoretical work above suggests that the cumulating cases weighting methodology may increase the effective sample sizes (i.e., reducing the design effects). However, this is not the same as actually showing reduced design effects and increased effective sample sizes. We have calculated design effects and effective sample sizes using weights calculated by both methods for the first three NLSY97 rounds. We used a representative set of 30 key variables, 13 of which were mean variables (e.g., family income), and 17 of which were proportions (e.g. is the youth a high school graduate?).

For each variable, we calculated the effective sample size using each of the two weights. Then, we calculated the relative increase or decrease in effective sample size using the cumulating cases weight rather than the combining samples weight. This is the ratio of the effective sample sizes minus one. Treating means and proportion variables separately, we then calculated the median increase/decrease in effective sample size for each round, overall, and by race/ethnicity group.

The results are shown in Table 5, along with the theoretical results discussed in Section 5 for all three rounds. Looking at the Round 2 Theoretical line, you can see that the 5.1% overall increase in effective sample size matches the data in Table 2 (7022/6705 = 1.051) indicating that the cumulating cases method increases the effective sample size by 5% of the total sample. The theoretical results are very consistent for all three rounds, and are surprisingly accurate predictors of the median gains among means and proportions. Table 5 confirms that the gain in effective sample size overall is 4-5%, or approximately 300 interviews. It also confirms that the percentage increases for the minority sub-samples are even greater: 20-30% for non-Hispanic Black youths, and 30% or more for Hispanic youths. Of course, the gain is 0% for non-Hispanic, non-Black youths because only minority youths were eligible in the SU sample.

7. Conclusions. Table 5 shows that the cumulating cases strategy is superior to the combining samples strategy, using the very same data. Essentially, the only additional "cost" associated with the cumulating cases strategy is the work needed in obtaining the selection probabilities for all cases for both samples. Planned in advance, this is likely to be very little work. However, it could be very difficult to cumulate across separate surveys, especially if different companies or people created the surveys.

A system that combines probabilities for the individual cases and then cumulates cases is conceptually superior to the sample-based system in section 3. In particular a case-based system ensures that cases with the same overall probabilities of selection will receive the same weights. Under the sample-based system this will not necessarily be the case. Consider, for example, two Hispanic males, each resident in the same primary sampling unit, each with the same probability of selection in the CX sample and with the same probability of selection in the SU sample, and therefore the same overall probability of selection. One might be selected into the sample as part of the CX sample, the other as part of the SU sample. Under the combined sample weights in Section 3, the weights allocated to the two individuals would be different; under the cumulated cases weighting scheme in Section 4, their weights would be the same. We argue that the latter is the more appropriate outcome. The cumulating cases strategy is preferable because it is conceptually superior, it is almost cost-free, and in our case, it resulted not only in a significant increase in efficiency for minority estimates, but also a modest increase in efficiency for overall estimates. We recommend using the *cumulating cases* weighting strategy rather than the *combining samples* weighting strategy, whenever appropriate, and note that starting with Round 4, NLSY97 weights are being calculated using the *cumulating cases* weighting strategy.

References:

Frankel, M.R., McWilliams, H.A., and Spencer, B.D. (1983), National Longitudinal Survey of Labor Force Behavior, Youth Survey (NLS) Technical Sampling Report. National Opinion Research Center.

Kish, L (1965) *Survey Sampling*. New York: John Wiley and Sons.

TABLE 1. NLSY97 SAMPLE SIZES BY SAMPLE AND RACE/ETHNICITY

Sample	Race/Ethnicity	Round 1	Round 2	Round 3
CX	Hispanic	1,079	1,019	991
	Non-Hispanic Black	923	852	836
	Non-Hispanic Non-Black	4,746	4,408	4,346
SU	Hispanic	1,254	1,183	1,139
	Non-Hispanic Black	982	924	897
BOTH	TOTAL	8,984	8,386	8,209

TABLE 2. OVERALL POPULATION ESTIMATES – ROUND 2

	CX only	Combining Samples	Cumulating Cases
n	6276	8386	8386
CV(w_i)	0.21	0.50	0.44
1+L	1.04	1.25	1.19
n_{eff}	6010	6705	7022
Gain vs CX only	n/a	695 from 2110 interviews	1012 from 2110 interviews
Gain: Cumulating vs Combining	n/a	n/a	317 [5% of total sample] [46% addition to SU] [no cost]

TABLE 3. ESTIMATES OF THE NON-HISPANIC BLACK POPULATION – ROUND 2

	CX only	Combining Samples	Cumulating Cases
n	1019	2202	2202 [1019CX+1183SU]
CV(w_i)	0.15	0.70	0.41
1+L	1.02	1.49	1.18
n_{eff}	995	1479	1872
Gain vs CX only	n/a	484 from 1183 interviews	877 from 1183 interviews
Gain: Cumulating vs Combining	n/a	n/a	393 [27% of non-Hisp. black sample] [81% addition to SU] [no cost]

TABLE 4. ESTIMATES FOR THE HISPANIC POPULATION – ROUND 2

	CX only	Combining Samples	Cumulating Cases
n	852	1776	1776 [852CX+924SU]
CV(w_i)	0.15	0.77	0.44
1+L	1.02	1.60	1.19
n_{eff}	833	1114	1490
Gain vs CX only	n/a	281 from 924 interviews	657 from 924 interviews
Gain:Cumulating vs Combining	n/a	n/a	376 [34% of Hispanic sample] [134% addition to SU] [no cost]

TABLE 5. REDUCTION IN DESIGN EFFECTS / % INCREASE IN EFFECTIVE SAMPLE SIZES

		Overall	Hispanic	Non-Hispanic Black	Other
Round 1	*Theoretical*	*5.1%*	*33.6%*	*27.2%*	*0.0%*
	Means	5.1%	28.5%	33.5%	0.0%
	Proportions	4.1%	19.8%	24.5%	0.0%
Round 2	*Theoretical*	*5.0%*	*33.9%*	*26.8%*	*0.0%*
	Means	3.6%	42.5%	17.8%	0.0%
	Proportions	4.4%	29.1%	21.3%	0.0%
Round 3	*Theoretical*	*5.1%*	*34.3%*	*26.4%*	*0.0%*
	Means	4.3%	40.7%	32.0%	0.0%
	Proportions	5.6%	34.1%	25.0%	0.0%