BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES
SRD Research Report Number: CENSUS/SRD/RR-86/23

WEIGHTING ISSUES FOR LONGITUDINAL HOUSEHOLD AND FAMILY ESTIMATES

by

Lawrence R. Ernst
Statistical Research Division
Bureau of the Census
Room 3524, F.O.B. #3
Washington, D.C. 20233 U.S.A.

Recommended by: Kirk M. Wolter

Report completed: October 24, 1986

Report issued: November 17, 1986 (Revised)

1. INTRODUCTION

In the last few years several surveys have been conducted which have had as one on their goals to tell us what happens to households and families over time. These include the National Medical Care Expenditure Survey (NMCES), the National Medical Care Utilization and Expenditure Survey (NMCUES) and the Survey of Income and Program Participation (SIPP). The process of obtaining longitudinal estimates for such units presents some important questions that either are not encountered or are easier to answer for cross-sectional estimation or longitudinal person estimation. The following are three key questions:

1. Since the composition of households and families can and does change over time, which changes should allow the unit to be considered still continuing and which mark the dissolution of the unit?

2. What operational rules should be used to determine which households, families and individuals are to be followed over time, and what retrospective questions should be asked of individuals who join the sample after the beginning of the panel?

3. What weighting procedures should be employed to obtain weights that yield unbiased estimates, and how should the weights be adjusted to reduce the variances and biases of the estimates?

In this paper the focus is on the third question, but as will become clear in Section 4, in order to obtain unbiased estimates, the right combination of weighting procedure, longitudinal household or family definition and operational procedures is required.

In Section 2 we state the assumptions that are made in this paper and fix notation and terminology. In Section 3 after explaining why a common type of weighting procedure used in sampling to obtain unbiased estimates, weighting by the reciprocal of the probability of selection, does not, in practice, generally work for longitudinal household and family estimation, a class of weighting procedures which can accomplish this task is presented. In Section 4, we explain the difficulties that can arise in actually applying these weighting procedures because the information necessary to determine the weight, the continuity of a household or family, or some of the subject-matter data needed in the estimates may not be collected under the assumed operational procedures. Also presented are conditions which, if satisfied by a longitudinal household or family definition, are sufficient for there to exist a weighting procedure that avoids these difficulties. Finally, in Section 5 some thoughts on the adjustment of the weights used to produce unbiased estimates are discussed. This discussion focuses on procedures for adjusting for noninterview and for controlling estimates for key demographic variables to independent estimates, since these procedures may have to be handled differently for longitudinal household and family estimation than for either cross-sectional estimation or longitudinal person estimation.

Many of the ideas in this paper were originally developed in Whitmore, Cox and Folsom (1982), Ernst (1983, 1985, 1986) and Ernst, Hubble and Judkins (1984). In particular, this author wishes to acknowledge the work of his co-authors in the last mentioned paper that have been incorporated into the present paper.

2. PRELIMINARIES

In order to keep the discussion in this paper from becoming overly complex, a fixed set of design and operational procedures is assumed throughout this paper and will now be described. The set of procedures chosen is motivated by the procedures used in NMCES, NMCUES and SIPP, which bear many resemblances. The presentation in this paper is, for the most part, with respect to households only, not families. Since a family is a subset of a household it should be relatively easy to make the necessary modifications for family estimation.

We take a month to be the basic unit of time. For each month t, C_t denotes a cross-sectional universe of households and

P_t the set of individuals residing in a household in C_t. For example, C_t might be the set of all households in the United States that contain civilian persons. The initial sample at month B is a probability sample of members of C_B. An individual in a chosen household is known as an <u>original sample person</u>. There are several rounds of interviewing with each interview covering the month or months since the previous interview; E denotes the final interview month for the sample panel. For each month t, the set of individuals to be interviewed are all original sample people in P_t plus all other people in P_t residing with an original sample person. This latter group of people are referred to as <u>associated sample people</u>. Note that associated sample people are interviewed only for months in which they reside with an original sample person. When associated sample people are initially interviewed they are asked only enough retrospective information to ascertain whether they were in P_B.

A longitudinal household (LHH) universe, L, is a collection of disjoint subsets of $\bigcup_{t=B}^{E} C_t$ such that each member H of L is of the form $H=\{A_b, A_{b+1}, \ldots, A_e\}$, where $A_t \in C_t$, and b, e are the beginning and ending months respectively of the LHH (the meaning that b and e have throughout this paper). The definition of a specific LHH universe consists of two parts. The first part is the LHH definition itself, which we consider a set of rules which for any $A_t \in C_t$ specify which $A_{t+1} \in C_{t+1}$, if any, is eligible to be in the same member of L. The following are six examples of such rules. Each of the first five rules constitutes a LHH definition by itself, or they can be used in combination (in some cases redundantly) in a LHH definition. The final rule does not constitute a LHH definition alone.

<u>Same Householder Rule (SH)</u>. A_t and A_{t+1} have the same householder. (The householder of a household, in Census Bureau terminology, is the first adult person listed on the household roster. According to the survey instructions this person should be an owner or renter of the housing unit.)

Same Spouses Rule (SS). If A_t is a married-couple household then A_{t+1} is also a married-couple household with the same husband and wife; otherwise A_t and A_{t+1} have the same householder.

No change Rule (NC). A_t and A_{t+1} have the same household members.

Reciprocal Majority Rule (RM). The majority of members of A_t are in A_{t+1} and the majority of members of A_{t+1} are in A_t.

Reciprocal Plurality Rule (RP). A_{t+1} contains more household members in A_t at month t than any other household in C_{t+1} and A_t contains more household members in A_{t+1} at month t+1 than any other household in C_t.

Household Type Rule (HT). Either A_t and A_{t+1} are both married-couple households, both other family households, or both nonfamily households.

NMCES and NMCUES essentially used the SS and RM rules respectively as LHH definitions. No LHH definition has been officially adopted yet for SIPP although a definition combining SS and HT is currently the leading candidate. It is not the purpose of this paper to compare any of these rules from a social science point of view. They will, however, serve to illustrate the discussion in Sections 3 and 4.

The second part of the definition of a specific LHH universe places restrictions on the form each member of L must take. For example, these might be the requirement that each LHH be in existence at least two months. Another example is the requirement that b=B and e=E, that is that the universe is restricted to LHHs in existence for the entire life of the panel. Finally, there may be the restriction that L consists of a cohort of LHHs in existence at month B, the initial LHHs, plus a set of LHHs formed after month B, the subsequently formed LHHs, "generated by" the set of initial LHHs. This last restriction only is assumed in this paper, for, as will be explained in Section 3, this assumption is necessary to obtain unbiased weights.

To illustrate the vague concept "generated by", if the LHH definition is SH then the set of subsequently formed LHHs generated by the initial LHHs might be all LHHs whose householder at month b is in P_B, while if the LHH definition is NC, it might be the set of LHHs with at least one member in P_B at month b, or, alternatively, the set of LHHs for which all members at month b are in P_B.

3. OBTAINING UNBIASED WEIGHTS

To motivate the approach to obtaining unbiased estimates for LHHs to be presented in this paper, it will first be explained why weighting by the reciprocal of the probability of selection is not in general feasible for this purpose, and hence the need to consider alternatives. Let $X=\sum_{i=1}^{N} x_i$ be a parameter of interest, where x_i is the value of the characteristics for the i-th unit in a population of size N. Typically in survey work to estimate X a sample would be drawn in such a manner that the i-th unit has a known positive probability of being chosen, and X would then be estimated by

$$\hat{X} = \sum_{i=1}^{N} w_i x_i , \qquad (3.1)$$

where

$$\begin{aligned} w_i &= \frac{1}{p_i} \quad \text{if the i-th unit is in sample,} \\ &= 0 \quad \text{otherwise.} \end{aligned} \qquad (3.2)$$

Unfortunately, for LHH estimation such an approach is generally not practical, as is illustrated by the following example. Consider any subsequently formed household under the NC definition. Such a LHH would be in sample if and only if at least one household member is an original sample person and, consequently, to use (3.1) and (3.2) as an estimator it would be necessary to determine the probability of this event. It would be operationally impossible to determine this probability, since it would first be necessary to determine the first round household for each member of the current household, including associated sample persons, and then compute the probability that at least one of these first round households was selected.

Fortunately though, it is not necessary that w_i satisfy (3.2) is order that (3.1) be unbiased. If fact, if w_i is any random variable associated with the i-th unit in the population satisfying

$$E(w_i) = 1, \tag{3.3}$$

then (3.1) is unbiased, that is $E(\hat{X}) = X$. Thus, defining unbiased LHH weighting procedures reduces to defining random variables w_i satisfying (3.3).

All of the weighting procedures that have been proposed for obtaining unbiased LHH estimates which satisfy (3.1) and (3.3) have been of the following form. Let M denote the number of individuals in P_B. Associate with the j-th individual in P_B, a weight, w_j', as follows. Let p_j denote the probability that the j-th individual's household is in sample at month B, and then let

$$\begin{aligned} w_j' &= \frac{1}{p_j} \text{ if the individual's household is in sample at month B,} \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{3.4}$$

Then for the i-th LHH in L associate a set of constants α_{ij}, $j=1,\ldots, M$, with α_{ij} independent of w_j and

$$\sum_j \alpha_{ij} = 1. \tag{3.5}$$

Finally, let the weight w_i of the i-th LHH be

$$w_i = \sum_j \alpha_{ij} w_j' . \tag{3.6}$$

Clearly, any set of w_i's which satisfy (3.4-3.6) also satisfies (3.3). Furthermore, and these are the crucial differences, while (3.2) in general requires knowledge of the joint probability of selection for some members of C_B, (3.4-3.6) only require knowledge of individual probabilities of selection, and while (3.2) requires that the probability of selection be known for some members of C_B not in the initial sample, (3.4-3.6) does not.

The most common examples of estimators satisfying (3.4-3.6) are those for which associated with the i-th LHH is a set S_i of household members such that

$$\alpha_{ij} = \frac{1}{m_i} \text{ if the j-th individual is in } S_i \cap P_B, \qquad (3.7)$$
$$= 0 \quad \text{otherwise,}$$

where m_i is the number of individuals in $S_i \cap P_B$. (3.6) is thus the arithmetical average of the weights of the individuals in $S_i \cap P_B$. Note that (3.7) requires that $S_i \cap P_B$ be nonempty for each LHH in L.

Below are five examples of weighting procedures that have appeared previously in the literature (Ernst 1983), (Ernst, Hubble and Judkins 1984), and which satisfy (3.4-3.6). The first three examples are of the form (3.7) and also yield weights which do not vary with the interval for which the estimates are made. The fourth and the fifth examples are provided to illustrate weighting procedures which lack the latter property and the former property respectively.

1. Beginning Date of Household Procedure (BH). S_i is the set of all household members at month b.

2. Householder Weight Procedure (HW). S_i is the singleton set consisting of the householder at b.

3. Average of Spouses Weights Procedure (AWS). S_i consists of the householder and spouse at b if the householder is a married-couple household at b; otherwise the householder at b is the only member of S_i.

4. Beginning Date of Interval Procedure (BI). S_i is the set of all household members at the beginning of the time interval of interest.

5. Average of Monthly Weights Procedure (AMW). For the i-th LHH let d_i be the number of months that the LHH contains at least

one member of P_B, and for each month t let m_{it} be the number of such individuals. For the j-th person in P_B let T_{ij} denote the set of months that this individual is in the i-th LHH. Then let

$$\alpha_{ij} = \frac{1}{d_i} \sum_{t \in T_{ij}} \frac{1}{m_{it}} .$$

Note that this yields the same weight as would be obtained by averaging the weights of the individuals in P_B who are in the LHH each month and then averaging the result over all months that the LHH contains a member of P_B.

Corresponding to each of these weighting procedures is a largest universe for which the procedure is defined. For any procedure of the form (3.7) it is all LHHs for which $S_i \cap P_B$ is nonempty. For example, for BH it is all LHHs which at b contains at least one member of P_B, while for HW it is all LHHs for which the householder at b is in P_B. For AMW it is all LHHs that contain a member of P_B for at least one month. Each weighting procedure can of course also be used for any smaller universe. These examples illustrate why it is necessary, if strictly unbiased weights are desired, to restrict the universe to a cohort of initial LHHs plus a set of subsequently formed LHHs generated by the initial LHHs, as was indicated in Section 2.

In the next section we will show that certain of the weighting procedures that have just been defined in combination with certain of the LHH definitions in Section 2 can overcome some practical problems in obtaining unbiased estimates that might otherwise occur.

4. OVERCOMING PROBLEMS ASSOCIATED WITH OBTAINING UNBIASED WEIGHTS

Although estimators obtained using weighting procedures satisfying (3.4-3.6) avoid the difficulties arising from the use of (3.2), some of the information needed to compute (3.1) would still not be available for many combinations of weighting procedures and LHH definitions because it would not be collected under the assumed operating procedures. (This is true even if it is assumed, as we shall in this section, that there is no nonresponse and that there is perfect frame coverage.) This unavailable information can result in the following three problems for a LHH.

Problem A. The weight associated with the LHH is not known because it depends on information about the LHH before it entered sample or after it left sample that was not collected.

Problem B. The LHH has a positive weight and yet subject-matter data is incomplete because the LHH existed before entering sample or after leaving sample.

Problem C. The LHH has a positive weight and yet its period of existence is unknown. This problem can be divided into the following two subproblems.

1. It cannot be determined whether the LHH existed before entering sample or after leaving sample.

2. For some month t it cannot be determined if a household $A_{t+1} \varepsilon C_{t+1}$ is the continuation of a household $A_t \varepsilon C_t$ even though both A_t and A_{t+1} are in sample.

Note that problem A is equivalent to not knowing some of the w_i's in (3.1); problem B is equivalent to not knowing some of the x_i's when $w_i > 0$; and problem C is equivalent to not knowing the set of units to be used in (3.1).

To illustrate these problems, consider the RP rule as the entire LHH definition, together with the BH weighting procedure. With this combination, problems A, B, C.1 and C.2 can all occur. For example, suppose three associated sample people g_3, g_4, g_5, who had not previously been interviewed, move in at month u with two original sample people g_1, g_2, who since month B had been living together in a two person LHH with weight w. These five people remain together until month v when the three associated sample people leave. Let A_t, $u \leq t < v$, denote the household at month t consisting of these five people. Then the household arrangements of these five people at months u-1 and v are not entirely known. For example, at u the following are three possible arrangements:

1. $A_{u-1,1} = \{g_1, g_2\}$, $A_{u-1,2} = \{g_3, g_4, g_5\}$.
2. $A_{u-1,1} = \{g_1, g_2\}$, $A_{u-1,2} = \{g_3\}$, $A_{u-1,3} = \{g_4\}$, $A_{u-1,4} = \{g_5\}$.
3. $A_{u-1,1} = \{g_1, g_2\}$, $A_{u-1,2} = \{g_3, g_4\}$, $A_{u-1,3} = \{g_5\}$.

For arrangement 1, A_u is the continuation of $A_{u-1,2}$, and this LHH with rare exceptions has weight 0, since there usually would have been no original sample people present when it began. For arrangement 2, A_u is the continuation of $A_{u-1,1}$ and the LHH has weight w. For arrangement 3, A_u is a member of a LHH which began at month u with weight 2w/5. Thus, clearly problem A occurs for the LHH H for which A_u is a member. Furthermore, if H has a positive weight, as we shall assume from now on, then problem C.2 occurs since it is not known if A_u is a continuation of $A_{u-1,1}$.

Now at month v the same three arrangements can occur with v replacing u-1. Then A_{v2} is the continuation of A_{v-1} in arrangement 1; A_{v1} is the continuation of A_{v-1} in arrangement 2; while for arrangement 3, month v-1 is the final month for H. Problems C.1 and C.2 both occur, as is illustrated by arrangements 1 and 2. Furthermore, for arrangement 1, problem B arises since $A_{v2} \in H$ but A_{v2} is not in sample.

Overcoming problems A, B, and C simultaneously for all LHHs with the current operational procedures requires an appropriate combination of LHH definition and weighting procedure. The key result of this section is that if the following two conditions are satisfied by the LHH definition, then there exists a weighting procedure which avoids all three problems.

Condition 1. For each month and each $A_t \in C_t$ there exists a nonempty subset A_t^* of A_t which depends only on A_t, such that $A_{t+1}^* = A_t^*$ is necessary for A_{t+1} to be the continuation of A_t.

Condition 2. The determination of whether $A_{t+1} \in C_{t+1}$ is the continuation of $A_t \in C_t$ depends only on A_t and A_{t+1}.

If conditions 1 and 2 are satisfied then, as we will shortly show, problems A, B and C are avoided if a weighting procedure of the form (3.7) is used with $S_i = A_b^*$. First, however, let us illustrate these conditions. Among the six LHH definition rules presented in section 2, SH, SS and NC alone or in combination with any of the other rules, satisfy condition 1. For example, condition 1 would be satisfied for each of these rules if A_t^* consists of the householder of A_t alone, and this A_t^* yields the HW procedure. For SS and NC, A_t^* can also taken to be the householder and spouse in a married-couple household, and the householder only otherwise, leading to the ASW procedure. Finally, $A_t^* = A_t$ can be used for NC, yielding the BH procedure. However, no combination of the other three rules satisfies condition 1.

Also note that for those LHH definitions for which more than one choice of A_t^* is available, it may be best to choose the weighting procedure corresponding to the largest possible A_t^*, since this would result in the largest number of LHHs with positive weights, which would in general result in an estimator (3.1) with the greatest precison. With the previous examples as an illustration, BH would be the choice for the NC definition, ASW for the SS definition, while HW is the only choice for SH.

As for condition 2, it is satisfied by any combination of these six rules not including RP. Condition 2 is not satisfied by RP alone. To see this, refer to the example presented earlier in this section in which whether A_u is a continuation of $A_{u-1,1}$ depends also on the composition at u-1 of households other than $A_{u-1,1}$. Note also that RP in combination with any of the other five rules except NC or RM still does not satisfy condition 2. However, RP is redundant in combination with NC or RM and condition 2 is then satisfied.

We next establish that if conditions 1 and 2 are satisfied then problems A-C are avoided with the stated weighting procedure. In fact, only condition 1 is required for all but Problem C.2. Clearly this is true for Problem A. Problem B is avoided since if a LHH $H=\{A_b,...,A_e\}$ has a positive weight then these exists an original sample person g that is a member of A_b^* and hence a member of A_t^* for all months t of H's existence, and therefore H is in sample throughout its period of existence. Similiarly, Problem C.1 is avoided since the LHH cannot be in existence before entering sample or after leaving sample, because g would not have been a member.

Condition 1 alone and this weighting procedure do not guarantee that Problem C.2 does not occur. For example, a definition combining SH and RP does not avoid this problem with the HW procedure. However, clearly condition 2 guarantees that Problem C.2 is avoided.

Finally, we remark that in general conditions 1 and 2 do not assure that complete information is available for each LHH that ever appears in sample. In fact, only the NC rule assures that. What these conditions together with the specfied weighting procedure does guarantee is that all LHHs that are not in sample for the entire period of existence are zero-weighted and consequently that the missing information is not needed.

5. WEIGHTING ADJUSTMENTS

To obtain final weights for use in producing estimates for household surveys, the unbiased weights are typically subjected to one or more adjustment procedures for the purpose of reducing the variance of the estimates produced and the bias resulting from undercoverage and nonresponse. For example, at the Census Bureau these adjustments generally include a noninterview adjustment, an adjustment for the purpose of reducing between primary sampling units variablity and an adjustment to independent estimates of key demographic characteristics of the analytic unit.

Adjustments of the weights for a sample of LHHs would generally incorporate the same basic concepts as adjustments for a cross-sectional sample, but present additional complications arising from the time element. In this section we outline one approach to weighting adjustments for LHHs. Other approaches are certainly possible.

Our approach is motivated in part by the nature of our assumed universe, that is a cohort of initial LHHs plus a set of subsequently formed LHHs generated by the initial LHHs. The procedure envisioned consists of the following three major steps.

<u>Step 1.</u> Adjust the weights of the set of initial sample LHHs much as one might adjust a cross-sectional sample at month B, through a series of adjustments culminating in an adjustment to independently derived estimates of the number of cross-sectional households with specific characteristics in existence at B.

<u>Step 2.</u> Next adjust the weights of the subsequently formed sample LHHs to carry over to these LHHs the adjustments to the initial sample LHHs. This includes adjusting for noninterviews among subsequently formed LHHs that result from noninterviewed initial sample LHHs.

<u>Step 3.</u> Finally, further adjust the set of subsequently formed sample LHHs to account for the other category of noninterviews among these LHHs, noninterviews which result from original sample people in interviewed initial sample LHHs who later become noninterviews.

Before detailing these steps, we digress to present observations on two general aspects of the adjustment which affect more than one of these steps. The first of these is the question of the number of points in time for which the sample estimates should be controlled to independent demographic estimates. The proposed procedure envisions doing this only at month B. One fundamental reason for not proposing adjustment at more than one point in time is that the LHH universe that we are considering excludes subsequently formed LHHs not generated by initial LHHs. Consequently, the number of LHHs in the universe at any time after month B would not agree with independent cross-sectional estimates which include such households. Furthermore, even if appropriate controls can be obtained at more than one point in time, there would be difficulties in attempting to obtain agreement with these controls. One approach to obtaining this agreement is to group the LHHs in each cell according to their pair of beginning and ending months and to then apply a different weighting factor to each such group. The values for these factors can be determined by considering them as variables in a mathematical programming problem. This approach is described in detail in Judkins et al. (1984). However, in certain situations no solution would be possible unless some weighting factors are allowed to be very large or negative, and sometimes not even then. For example, if it is desired to control to independent estimates at two months t_1 and t_2 and for some cells the indentical set of sample units are in the cell at both months while the control totals are different, then agreement cannot be reached.

The second item of discussion is whether the final weights for LHHs should vary with the interval for which estimates are produced, even when the unbiased weights do not. This is also an issue in longitudinal person estimation, and in fact was first discussed by Kobilarcik (1985). It arises in the noninterview adjustment problem as a result of the fact that there can be a considerable number of sample LHHs which are interviewed for some but not all of their period of existence. If one final weight is used for each LHH, then such LHHs would have to be considered noninterviews in the noninterview adjustment process (unless data were imputed for the missing time periods) and the data collected from these LHHs would not be used directly in the estimation. In contrast, the use of final weights that vary with the time interval for which estimates are to be made, allow the use of LHHs that are interviewed for some rounds, but not all, to be used in estimates for time intervals throughout which they are interviewed. This would be accomplished by allowing the set of LHHs considered noninterviews in the noninterview adjustment process to vary with the time interval, which would cause the noninterview adjustment factors and final weights to vary also.

A drawback to the use of more than one final weight for a LHH is that this increases the operation complexity. Even if more than one final weight is produced, the number of different intervals for which distinct final weights are obtained may have to be limited to keep the processing problem from becoming unmanageable. To simplify this problem, it will be assumed that the noninterview pattern for each LHH and person is nested, that is noninterview for one month implies noninterview for all subsequent months (Then for any actual case for which the noninterview pattern is not nested, either missing interviews would be imputed or interview data subsequent to the first noninterview month would not be used in the estimation.) It would then be appropriate to obtain weights for a limited number of intervals of the form [B, t_i], i=1,...,k say. Then if estimates are desired for an interval [t, t′], the weights associated with the smallest of the intervals [B, t_i] containing [t, t′] would be used in the estimation.

We now detail the suggested three steps in the adjustment process. It is assumed in this discussion that a single final weight for the interval [B, E] is to obtained for each sample LHH. However, by simply considering t as the final month of the panel, this process would apply equal well for any interval of the form [B, t], and hence can be also used if the final weights vary with the interval for which estimates are to be produced.

1. Adjustment of Weights Initial Sample LHHs. Conceptually, the adjustment procedures for initial sample LHHs are similar to the adjustments for a cross-sectional survey at month B, and we consequently highlight here only aspects for which there may be important differences. Noninterview adjustment is one such area. At the Census Bureau, for example, a single noninterview adjustment is generally used in the household surveys for cross-sectional estimation. In this adjustment the analytic units are partitioned into adjustment cells defined by demographic characteristics of the unit. The weight of each interviewed unit is multiplied by a noninterview adjustment factor, namely, the sum of the weights of the interviewed plus the noninterviewed units in the cell divided by the sum of the weights of the interviewed units; and the noninterviewed units are then zero-weighted, thereby redistributing the weights of the noninterviewed units in each adjustment cell to the interviewed units. For the initial sample LHHs, however, it is proposed that the noninterview adjustment be performed in two steps. In the first step the weights of units not interviewed at B would be redistributed to all other units in the cell, while in the

second, the weights after the first adjustment of units interviewed at B but not for their entire period of existence would be redistributed to the initial LHHs interviewed for their entire period of existence. The reason for proposing two such adjustments here is that this would permit a selection of variables to use in forming adjustment cells from the extensive data collected in previous interviews from those LHHs interviewed at B, instead of being restricted to the limited information that typically would be available for LHHs not interviewed at all. (The concept of using two noninterview adjustments in this context applies equally well to longitudinal person estimation, and in fact was first developed for SIPP longitudinal person estimation (Jones 1986).)

The first noninterview adjustment presents no unusual difficulties, and in fact can be done precisely as a noninterview adjustment at B in a cross-sectional survey. However, there are at least two complications that arise in the second adjustment. First, if a LHH had month t as its first noninterview month, then it might not be known whether the LHH actually continued to exist at month t. This information is important since such a LHH can obviously only be a noninterviewed LHH if it continued to exist at month t. Imputation may be necessary to make this determination. Secondly, it may be desirable to redistribute the weights of LHHs with first noninterview month t only to interviewed LHHs still in existence at month t by computing noninterview factors F_{tc} that vary not only with the adjustment cell C, but also with t. To compute F_{tC}, first let I_{tC} denote the weighted count in cell C (using the weights after the first noninterview adjustment) of interviewed LHHs with period of existence [B, t] and let N_{tC} denote the weighted count of noninterviewed LHHs in cell C with first noninterview month t. (Note that $N_{BC} = 0$ because of the first noninterview adjustment.) Then let

$$F_{tC} = 1 + \sum_{i=B}^{t} \left(\frac{N_{iC}}{\sum_{j=i}^{E} I_{jC}} \right).$$

Application of this factor redistributes the weights of all noninterviewed LHHs in cell C with first noninterview month t to all interviewed LHHs in existence at month t. Furthermore, the sum of the weights of all interviewed LHHs in cell C after this adjustment is

$$\sum_{t=B}^{E} F_{tC} I_{tC} = \sum_{t=B}^{E} \left[1 + \sum_{i=B}^{t} \left(\frac{N_{iC}}{\sum_{j=i}^{E} I_{jC}} \right)\right] I_{tC}$$

$$= \sum_{t=B}^{E} I_{tC} + \sum_{i=B}^{E} \left[\frac{N_{iC} \left(\sum_{t=i}^{E} I_{tC} \right)}{\sum_{j=i}^{E} I_{jC}} \right]$$

$$= \sum_{t=B}^{E} I_{tC} + \sum_{i=B}^{E} N_{iC},$$

which as desired is the sum of the weights before this adjustment of all LHHs in cell C, both interviewed and noninterviewed.

Another important question that arise in the weighting adjustment for initial LHHs is what to use as a source for the independent controls. Among U.S. national surveys, the Current Population Survey (CPS) estimates have been the choice in NMCUES, and also for SIPP longitudinal person and cross-sectional estimation. It should be noted that until recently CPS household and family estimates appropriate to use as controls were produced only for March estimates in conjunction with the Annual Demographic Supplement, and hence this was the month for which NMCUES estimates were controlled. Since the advent of SIPP these estimates have been produced monthly to provide controls for SIPP cross-sectional estimates, and they can also used as controls for LHH estimates.

Some necessary imperfections in the CPS household control totals should be noted. Although CPS estimates of the total number of persons in a given age-race-sex category are themselves controlled to independent demographic estimates which have no sampling variability, there are no such controls for household estimates. Consequently, such key CPS household estimates as number of households with householder in a specific age-race-sex category, or number of households of a given size or type, or even total number of households are subject to sampling variability and unknown biases. Despite this drawback, it is felt that adjusting to CPS estimates would be worthwhile in reducing sampling variablity and many biases because of the large size of the CPS sample and the relative reliability of CPS estimates.

As for the specific variables to use in the control process, this would of course depend on the needs of the particular survey. The variables used in NMCUES are presented in Whitmore, Cox and Folsom (1982).

2. Carry-over of Weight Adjustments for Initial Sample LHHs to Subsequently Formed Sample LHHs. Whatever undercoverage affects the set of initial LHHs also affects the subsequently formed LHHs generated by them. To compensate for this in a simple fashion, the weights for the set of sample subsequently formed LHHs can be adjusted by modifying (3.4) so that the final weight of each original sample person's initial LHH is used in place of the reciprocal of probability of selection. The motivation for this adjustment is that for a subsequently formed sample LHH which contains original sample people from a single initial sample LHH, as is nearly always the case, the ratio of the weight after this adjustment to the unbiased weight would appropriately be the same as the ratio of the final weight to the unbiased weight for the corresponding initial sample LHH.

3. Additional Noninterview Adjustments for Subsequently Formed Sample LHHs. Even if all original sample people that were members of interviewed initial sample LHHs continue to be interviewed throughout the life of the panel, there would still generally be for each $t \in (B,E]$, a set of noninterviewed LHHs formed at month t resulting from the noninterviewed initially formed LHHs. This set of noninterviews is compensated for by the

adjustment in step 2. However, in practice, there is also a set of noninterviewed LHHs formed at month t, denoted N_t', whose noninterview status results from later noninterviews among original sample people that were members of interviewed initial sample LHHs. Additional noninterview adjustments are required to compensate for this latter set. These adjustments present some significant complications that would not be found in longitudinal person estimation for example. To illustrate, consider the case of a sample initial LHH with final weight w that moves at month t and is not followed. Prior to the move the LHH contained five people but no information is available concerning the composition after the move. Then, at one extreme each of these five people might be living alone at month t, in which case the initial LHH generates five new LHHs at month t. At the other extreme these five people might remain together, in which case there are no new LHHs at month t generated by the initial LHH. Furthermore, the weight of any new LHHs would in general not be known. For example, with the ASW procedure, if one of these people is living alone at month t, the weight of this newly formed LHH after step 2 would be w. However, if that person instead forms a two person LHH by marrying an associated sample person, the corresponding weight would be w/2. Finally, if the person becomes part of a LHH in which the householder and spouse, if present, are associated sample persons, then the LHH would be zero-weighted. Thus, in addition to the problem of missing subject-matter data, noninterviewed LHHs in N_t' entail the additional problems of determining the number of noninterviewed analytic units and their weights. It is envisioned that these problems would have to be handled by some form of imputation procedure.

Once this imputation is performed, it is proposed that the LHH weights for the set of sample subsequently formed LHHs be adjusted to compensate for noninterviews in N_t', for any t>B, through a sequence of noninterview adjustments computed by using recursion on t as follows. First partition N_t' into two subsets, N_t'' and N_t''', the former consisting of noninterviewed LHHs which would have been interviewed had all original sample people interviewed at month t-1 continued to be interviewed throughout the life of the panel, and the latter consisting of the remaining LHHs in N_t'. Noninterviewed LHHs in N_t'' can be adjusted for by redistributing their weights to interviewed LHHs. However, the members of N_t''' would be unknown and must be compensated for by prior noninterview adjustments. Since any member of N_t''' must have resulted from noninterviewed LHHs that were in N_i'' for some i<t, the noninterviewed LHHs in N_t''' can be adjusted for by carrying over the adjustments for N_i'' for each i to LHHs formed after month i.

The following is the suggested noninterview adjustment procedure to compensate for noninterviews in N_t''. For each month t>B, a noninterview adjustment factor f_{tH} would be applied to each member H of the set of interviewed LHHs formed at t, denoted I_t, as follows. For each month $i \in (B,t)$ any $H_i \in I_i$ would have previously received a noninterview adjustment factor f_{iH_i} to compensate for noninterview in N_i'. From these factors, each original sample person interviewed at month t-1 would be assigned a month t-1 adjusted weight

$$w \prod_{i=B}^{t-1} g_{iH_i}, \qquad (5.1)$$

where w is the final weight of the person's initial LHH, H_i is the person's LHH for month i and

$$\begin{aligned} g_i &= f_{iH_i} \text{ if } H_i \text{ was formed at month i,} \\ &= 1 \text{ otherwise.} \end{aligned}$$

Thus, a noninterview adjustment factor would be applied to each original sample person for each month after B that the person was a member of a newly formed interviewed LHH. Then to compute f_{tH}, first compute an adjusted LHH weight w_{tH*} (but not the final weight) for each LHH $H^* \in I_t \cup N_t''$, using (3.6) and (3.4), with (5.1) substituted for the reciprocal of the probability of selection of the person's initial LHH. This is where the recursion occurs f_{tH} is then the sum of the weights w_{tH*} of all LHHs $H^* \epsilon I_t \cup N_t''$ -in the same adjustment cell as H divided by the sum of the same weights of all LHHs only in I_t in this adjustment cell. The final LHH weight for H would then be the product $f_{tH}w_{th}$. Note that this final weight is also the weight that would be obtained from (3.6) and (3.4) with each person's month t adjusted weight replacing the reciprocal of the probability of selection of the person's initial LHH. Thus, the final weight for H includes the factor f_{tH} to compensate for noninterviewed LHHs in N_t'', factors f_{iH_i} to account for noninterviews in N_t''' that result from noninterviews in N_i'', and the weight after step 2 which accounts for noninterviewed LHHs formed at month t resulting from noninterviewed initial LHHs.

In practice, there would be at least one major difficulty, in addition to its general complexity, in computing the f_{tH} factors using the method just outlined. The set I_t for a fixed t may often be too small to form adjustment cells containing a sufficient number of cases. Consequently some compromise may be necessary to the principal that members of N_t'' should have their weights distributed only to members of I_t.

REFERENCES

Cox, Brenda G., and Cohen, Steven B. (1985), Methodological Issues for Health Care Surveys, New York: Marcel Dekker.

Ernst, Lawrence R., (1983) "Preliminary Thoughts on Unbiased Longitudinal Household Estimation," U.S. Bureau of the Census report.

________(1985), "The Impact of Conditions in a Longitudinal Household Definition on the Ability to Produce Unbiased Longitudinal Estimates in SIPP," Report CENSUS/SRD/RR-85/16, U.S. Bureau of the Census, Statistical Research Division.

________(1986), "SIPP Longitudinal Household Estimation for the Proposed Longitudinal Household Definition," Report CENSUS/SRD/RR-86/5, U.S. Bureau of the Census, Statistical Research Division.

________, Hubble, David L., and Judkins, David R., "Longitudinal Family and Household Estimation in SIPP," Proceedings of the Section on Survey Research Methods, American Statistical Association. pp. 682-687.

Hanson, Robert H. (1978), The Current Population Survey - Design and Methodology, U.S. Bureau of the Census, Technical Paper 40.

Jones, Charles D. (1986), "SIPP 1984- Specifications for Longitudinal Weighting of Persons," U.S. Bureau of the Census memorandum to Thomas C. Walsh, July 17. Prepared by Edward Kobilarcik.

Judkins, D. R., Hubble, D. L., Dorsch, J. A., McMillen, D. B. and Ernst, L. R. (1984), "Weighting of Persons for SIPP Longitudinal Tabulations," Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 676-681.

Kobilarcik, Edward (1985), "Overview of SIPP Longitudinal Weighting," U.S. Bureau of the Census report.

Whitmore, Roy W., Cox, Brenda G., and Folsom, Ralph E. (1982), "Family Unit Weighting Methodology for the National Household Survey Component of the National Medical Care Utilization and Expenditure Survey," Research Triangle Institute report.