

## 1. INTRODUCTION

Many types of statistics will be produced by the Survey of Income and Program Participation (SIPP), but there is one type that was the driving force behind the unique design of the survey. To be fully successful, SIPP must tell us what happens to households over the course of time. From it we must obtain estimates of the patterns of income receipt, program participation, and labor force participation at the household and family level by a host of other characteristics. Of particular interest are parameters such as total annual household income and the number of families that have stopped drawing food stamps by demographic characteristics.

Before estimates can be produced, a decision must be made on the definition of a longitudinal household to be used in this survey. (To simplify the presentation, we will concentrate our discussion on longitudinal households as opposed to longitudinal families. However, parallel longitudinal estimation procedures can readily be developed for families). It often happens that the occupants of several housing units move and regroup. We need to know which, if any, of the resulting households are to be considered continuations of the previous households. Many definitions have been proposed, but final agreement has thus far not been achieved. Also decisions have yet to be made on whether households that form or dissolve during a time interval of interest are to be considered as part of the universe for estimation purposes. Because of the absence of agreement in these areas, several proposed definition and universe combinations will be considered in this paper. They are listed in Section 2. Also because of this absence of agreement, the major aim of this paper will be simply to compare several possible longitudinal household estimation procedures and present criteria for choosing among them, without attempting to reach a conclusion on a preferred procedure.

We foresee several steps in the process of producing longitudinal households estimates. The focus in this paper is the first step, the production of weights that would yield unbiased estimates assuming there are no data that are missing or in error, and that the frame coverage is perfect. Several procedures for obtaining such weights will be presented in Section 3. Choosing among these procedures is complicated by the fact that even assuming perfect response, data needed to produce unbiased estimates will be missing for some households because they are not collected with the current field procedures. This difficulty is principally due to the fact that, except for a few household definitions, all unbiased procedures assign positive weights to some longitudinal households for time periods when they are not in sample. The severity of this problem and the extent to which it is correctable in the future by changing field procedures or by modeling the missing data, vary by procedure. This problem, along with

descriptions of other important features, both positive and negative, that estimation procedures may possess is presented in Section 4. Finally, in Section 5 a detailed comparison of the features of the estimation procedures under consideration in this paper is presented.

It is assumed in this paper the reader has a basic knowledge of SIPP, including the design of this survey. Nelson, McMillen, and Kasprzyk (1984) provides this information.

Portions of the original paper, principally an examples section and a section on adjustments to the unbiased weights, are omitted here due to lack of space. The complete paper is available from the authors.

## 2. LONGITUDINAL HOUSEHOLD DEFINITIONS

In this section three possible longitudinal household definitions are presented to illustrate the longitudinal weighting procedures that will be described in the next section. (A fourth definition, known as the Shared Experience Definition was included in the original paper, but omitted here due to lack of space. In terms of the properties discussed in Section 5 it is identical to the Reciprocal Majority Definition that is included here.) A thorough discussion of longitudinal household definitions is presented in McMillen and Herriot (1984). In addition, several other terms will be defined, including the longitudinal household universes considered in this paper.

Since household composition and data for SIPP are obtained on a monthly basis, each of the definitions to be presented will be in terms of household continuity from one month to the following month. A longitudinal household over a time interval of  $n$  ( $\geq 2$ ) months is then defined to be one which is continuous for each of the  $n-1$  corresponding pairs of consecutive months. (It has not yet been decided if this approach will actually be used in SIPP.)

For each of the definitions below the conditions for which household B at month  $t+1$  is the continuation of household A at month  $t$  are stated. One condition that we require that all the definitions share is that A and B are either both family households or both non-family households. The other conditions are: No Change Definition (NC). A and B have the same household members.

Same Householder Definition (SH). A and B have the same householder. As an alternative, householder could be replaced by principal person in this definition without altering any of the statements made about it in subsequent sections, provided the final estimation procedure in Section 3 is also modified accordingly. (The householder of a household is, roughly, the person who owns or rents the housing unit. The principal person is the wife in a married-couple household, and the householder in all other households.)

Reciprocal Majority Definition (RM). The majority of individuals who are both household members of A at time  $t$  and in the universe at time  $t+1$  are members of B at time  $t+1$ , and the

majority of individuals who are both household members of B at time t+1 and in the universe at time t are members of A at time t. (This type of longitudinal definition was originally developed by Dicker and Casady (1982) for use in the National Medical Care Utilization and Expenditure Survey (NMCUES).)

We will now clarify several other terms.

A household is said to be in existence over a time interval of  $n \geq 2$  months if it is longitudinal over that time interval. Its period of existence is the longest such time interval. In the case of a household which is defined cross-sectionally for a month t, but is not longitudinal over either of the two month intervals containing t, then the period of existence of the household is defined to be one month.

If  $t_1$  and  $t_2$  are any pair of months, and longitudinal estimates are to be made over the interval  $[t_1, t_2]$ , then the following two possibilities will be considered in subsequent sections for the universe of households for which estimates will be produced.

Restricted Universe(R). The set of all households in existence over the entire interval  $[t_1, t_2]$ .

Unrestricted Universe(U). The set of all household in existence for one or more months in  $[t_1, t_2]$ .

Each sample panel is interviewed eight times. Each of the eight rounds of interviews takes four consecutive months to complete and is known as a wave.

Finally, we define an original sample person to be a person that was in sample during the first wave and will be at least 15 years of age by the end of the panel.

### 3. UNBIASED WEIGHTING PROCEDURES

In this section we present five weighting procedures for computing estimates of totals or proportions for longitudinal households that would be unbiased in the sense that the expected value of the estimator over all possible samples is the parameter of interest assuming no data are missing or in error, and perfect frame coverage. Modifications and adjustments of these estimation procedures necessary because of the unrealistic nature of these assumptions are considered in the original paper, but are omitted here due to lack of space. Except for the Continuous Household Members procedure, which will only be applied to the restricted universe, all the procedures will be stated for the unrestricted universe. To apply them to the restricted universe simply zero weight each household which is not in continuous existence over the time interval of interest. Furthermore, unless otherwise stated, all the procedures will be applied to all four longitudinal definitions defined in Section 2.

First we will explain why a common method of estimation, weighting by the reciprocal of the probability of selection is not feasible for our purposes, and hence the need to consider

alternative procedures. Let  $X = \sum_{i=1}^N x_i$  be a parameter of interest, where  $x_i$  is the value of

the characteristic for 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  $p_i$  of being chosen, and X would then

$$\text{be estimated by } \hat{X} = \sum_{i=1}^N w_i x_i, \quad (3.1)$$

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

Unfortunately for household and family estimation in SIPP, both cross-sectionally and longitudinally, such an estimation approach is not practical. For example, cross-sectionally a household is interviewed and used in the estimation process for a given month if and only if at least one household member is an original sample person. Consequently, to use (3.1) and (3.2) as an estimator it would be necessary to determine the probability that at least one member of the current household is an original sample person. It would be operationally impossible to determine this probability, since it would first be necessary to determine the first wave households for all current household members and then compute the probability that at least one of these first wave households was selected.

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

$$E(w_i) = 1, \quad (3.3)$$

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

Before we present the longitudinal weighting procedures we will state what, for purposes of this paper, a cross-sectional household weight is, since most of longitudinal weighting procedures will be defined in terms of cross-sectional weights. The first wave cross-sectional weight for a sample household is taken here to be the reciprocal of the probability of selection. For all nonsample households in the universe this weight is defined to be zero. For any month after the first wave a different definition is necessary because of possible changes in household composition. So, the cross-sectional household weight for any such month is defined to be the mean of the first wave cross-sectional household weights for all persons in the household that month who will be at least 15 years of age by the end of the panel and who were in the universe during the first wave. This type of weighting procedure is currently being used in SIPP to produce cross-sectional estimates, hence the name. It is readily verifiable that the weights satisfy (3.3).

We also will leave it to the reader to verify that the weights for each of the longitudinal procedures to be presented satisfy (3.3) and

hence lead to unbiased estimators.

Beginning Date of Household Procedure (BH). Each longitudinal household receives a single weight valid for any time interval that contains at least part of the period for which the household existed, namely the cross-sectional weight for the household at the beginning date of the household. In particular, if there were no original sample persons in a household at its beginning date then its longitudinal weight would be zero. This approach to longitudinal household estimation was previously used in the NMCUES (Whitmore, Cox and Folsom 1982).

Beginning Date of Time Interval Procedure (BI). Each longitudinal household receives a longitudinal weight valid for all time intervals with the same beginning date, namely the cross-sectional weight for the household at the beginning date of the time interval. Longitudinal households that form during the time interval are assigned the cross-sectional weight for the household at its beginning date, as in the preceding procedure.

Continuous Household Members Procedure (CM). The following procedure will only be applied to the restricted universe, as defined in Section 2. For any time interval for which the household is in existence the longitudinal weight to be assigned is determined by the set of persons that are members of the household throughout the time interval. The longitudinal household weight is the cross-sectional weight that would be assigned to a household consisting of this set of persons; that is, the average of the first wave weights of these people. A longitudinal weight of zero is assigned to the household if there are no original sample persons who are members throughout the time interval. The procedure is slightly biased because a longitudinal household with no members continuously present throughout a time interval has no chance of receiving a positive weight, thereby making satisfaction of (3.3) impossible. Since we believe this situation will rarely occur, at least for the longitudinal household definitions considered here, we expect this bias to be very small.

Average Cross-Sectional Household Weight Procedure (AW). Each longitudinal household receives a longitudinal weight valid for a specific time interval, namely the average of the monthly cross-sectional weights for the household over the intersection of the life of the household and the specified time interval.

Note, there are many procedures, like AW, that entail the averaging of weights, both household cross-sectional weights and person longitudinal weights. We will examine only one of these procedures here, as an example of this type of longitudinal household weighting procedure.

Householder Weight Procedure (HW). The following procedure will be applied only to the No Change and Same Householder Definitions, since it is appropriate only for definitions that allow for a single householder during the household's existence. (Generalizations of this procedure which are not so restricted in their applicability exist but will not be considered here.) The procedure assigns a single weight

valid for any time interval that contains at least part of the period for which the household existed, namely the first wave cross-sectional household weight of the householder's first wave household. A longitudinal weight of zero is assigned to the household if the householder was not an original sample person.

As will be seen in Section 5, this procedure is clearly the one of choice when the Same Householder Definition is used. If that type of definition is used with householder replaced by principal person then a similar modification of this estimation procedure with householder replaced by principal person would be appropriate.

#### 4. POTENTIAL ADVANTAGES AND DISADVANTAGES

The ideal unbiased weighting procedure would provide a single set of weights applicable to any time interval, require no more data than were collected, and possess the minimum variance among all unbiased procedures. Unfortunately, no such procedure exists. The procedures described in Section 3 all fail one or more of these three criteria to various degrees. In this section, we explain the nature of the failures without explicitly comparing the procedures. That is done in Section 5.

Multiplicity of Weights. Some procedures have the advantage of assigning to each household a single weight which depends only on conditions as of the first reference month for the household and which is valid for every interval that the household is in the universe. Other procedures have the disadvantage of sometimes producing different weights for the same household for different time intervals. (Procedures with this disadvantage could be modified so that only a single weight applies to any time interval, by computing for each household the weight appropriate for that procedure for the unrestricted universe and the 2 1/2 year time interval corresponding to the life of the panel. The weight obtained would also be used for any smaller subinterval for which the household is in the universe. However, weights obtained in this manner might not be able to be determined until the end of the life of the panel. This would make them difficult to use because we would have to wait until the last data from the panel were processed before estimates could be produced for any earlier time period. In any case, such weights would often lead to higher variances for short time intervals than weights developed specifically for the short time intervals.)

Unavailable Data Requirements. Most definition and procedure combinations require data from some households for time periods when the household is in existence but not in sample, that is for time periods for which interviews are not conducted for the household because no original sample people are members of the household. This needed data could be information for determining proper longitudinal weights or subject-matter information for use in tabulating the estimates. Some of this information is not collected for the 1984 panel of SIPP because of the current operational procedures. This is a consequence of the fact that agreement has not been reached on the longitudinal household definition to be used

in SIPP. In this vacuum, operational procedures were determined mainly by considerations of difficulty and cost. Once a definition has been agreed on, depending on the nature of the unavailable data, it might be possible to change operational procedures for future SIPP panels so that the required data are collected. To understand the problem with current operational procedures, consider the following situation. A household is longitudinal from month  $t_B$  to  $t_E$ . Original sample people are part of the longitudinal household only from month  $t_1$  to  $t_2$ . If  $t_B < t_1$ , then some prior information may be unavailable. Revised operational procedures to obtain this information might involve retrospective questions, longer reference periods or proxy data on anyone who left the household before the first interview. If  $t_2 < t_E$ , then some posterior information may be unavailable. Revised operational procedures might involve interviewing the household through  $t_E$ .

One of the important discriminants between the weighting procedures is how successfully they avoid the need for data from the period that the longitudinal household exists but is not in sample. (The need for such data is avoided by assigning zero weights to these problem households.) In terms of information needed for weighting, some procedures require only enough data to determine whether  $t_B < t_1$ , while others need to know  $t_B$  even when it is less than  $t_1$ . Similarly, some procedures only require knowledge of whether  $t_2 < t_E$ , while others need to know  $t_E$  even when it is greater than  $t_2$ . Furthermore, besides this need for information for determination of weights, if any parameters other than the number of longitudinal households are to be estimated, then required subject-matter data may be missing as well, either before  $t_1$ , after  $t_2$ , or both.

While the problem of missing information is a serious one, it is not fatal. Procedures can be developed to compensate for the unavailable data. Specifically, the data collected on these households while they were in sample should be sufficient for performing imputation for existence/nonexistence outside the in-sample period and formation and/or dissolution dates. The imputed values can then be used to calculate weights for these households. These households can then be treated as noninterviews so that the weights of mover households with similar demographic characteristics but with complete data receive increased weights while the deficient households themselves receives zero weights.

If the models underlying the procedures developed for adjusting for the missing information are true then it is still possible to obtain unbiased estimators, although now in a model-based sense. Furthermore, since the missing information that we are concerned with here is not caused by refusal to respond, modeling in this context might not suffer from the usually imperfect assumptions on similarity between respondents and nonrespondents that underlie any adjustments that use data from respondents to account for data missing from refusals. In addition, because of the longitudinal nature of the survey, there is gener-

ally a large amount of data available from the problem households that could be used in such adjustments. However, if the models are not perfect, then in general, the larger the proportion of data required that is unavailable, the greater the potential for serious bias problems.

Variations. In general, estimation procedures with the smallest variances are those that utilize available data intensively and tailor the weights to the specific time interval of interest. Unfortunately, as shall be seen in the next section, such procedures are often characterized by heavy needs for unavailable data which, as noted above, may impact unfavorably upon bias. Thus, there often is a direct trade-off between variance and the risk of bias. It will be difficult to weigh these factors against each other, since it appears that no single procedure will provide the correct balance for all of the multitude of characteristics that will be estimated by SIPP.

For use in the next section, we will define some labels for the advantages and disadvantages identified in the foregoing discussion. Let:

- $T_1$  mean that a single longitudinal weight exists for each household, valid for all time intervals for which the household is in the universe, and which depends only on conditions which could be determined during the first interview,
- $T_2$  mean the negation of  $T_1$ ,
- $BW_1$  mean that no data from the period preceding the first interview are unavailable but required for weighting,
- $BW_2$  mean that we need to know for weighting whether the longitudinal household existed before the first interview,
- $BW_3$  mean that we need to know for weighting the conception date of the household (within the time interval of interest),
- $BD_1$  mean that no subject-matter data from the period preceding the first interview are unavailable but required,
- $BD_2$  mean the negation of  $BD_1$ ,
- $FW_1$  mean that no data from the period following the last interview are unavailable but required for weighting,
- $FW_2$  mean that we need to know for weighting the dissolution date of the household (within the time interval of interest),
- $FD_1$  mean that no subject-matter data from the period following the last interview are unavailable but required,
- $FD_2$  mean the negation of  $FD_1$ .

Note that  $T_1$ ,  $BW_1$ ,  $BD_1$ ,  $FW_1$  and  $FD_1$  are the desirable properties.

##### 5. DETAILED COMPARISONS OF ADVANTAGES AND DISADVANTAGES

Table 1 below presents advantages and disadvantages of each definition, procedure and universe combination. A comparison of these features follows the table. Next, an explanation of each entry in the table is given. Finally, a discussion of data utilization, which is not in Table 1, is presented.

Table 1.  
Features

Defini- tion	Proce- dures	Uni- verse	T <sub>1</sub>	T <sub>2</sub>	BW <sub>1</sub>	BW <sub>2</sub>	BW <sub>3</sub>	BD <sub>1</sub>	BD <sub>2</sub>	FW <sub>1</sub>	FW <sub>2</sub>	FD <sub>1</sub>	FD <sub>2</sub>
NC	All	Both	X	X				X		X		X	
SH	HW	Both	X	X				X		X		X	
SH, RM	BH	U	X		X			X		X			X
SH, RM	BH	R	X					X			X		
SH, RM	BI	U		X	X			X		X			X
SH, RM	BI	R		X	X			X			X		X
SH, RM	CM	R		X	X			X		X		X	
SH, RM	AW	Both		X		X			X		X		X

Comparison of Features in Table 1. As noted at the end of Section 4, T<sub>1</sub>, BW<sub>1</sub>, BD<sub>1</sub>, FW<sub>1</sub>, and FD<sub>1</sub> are the desirable properties. For the NC definition all five procedures considered here possess all these desirable properties, as does the HW procedure for the SH definition.

However, for the SH and RM definitions, and most other definitions too, the BH, BI, and CM procedures have different subsets of the set of desirable features, so that the procedure to be adopted depends, at least in part on the features deemed most important. AW possesses none of these desirable features for these two definitions. Its principal advantage lies in possible reductions in variances because of complete utilization of available data, which will be discussed later. BH has advantages T<sub>1</sub>, BD<sub>1</sub>, and FW<sub>1</sub> for the unrestricted universe, and T<sub>1</sub> and BD<sub>1</sub> for the restricted universe. The main reason for consideration of this procedure would be that it is the only one among BH, BI and CM that always has advantage T<sub>1</sub>. BI has advantages BD<sub>1</sub> and FW<sub>1</sub> for the unrestricted universe and BW<sub>1</sub> and BD<sub>1</sub> for the restricted universe. Its principal advantage over BH is that for the restricted universe no retrospective questions need be asked. CM (which is only applicable to the restricted universe) possesses all desirable features except T<sub>1</sub>, that is no information not currently collected is needed for this procedure. Recall, however, that CM had the disadvantage of being slightly biased as explained in Section 3.

Explanation of Entries in Table 1. All explanations presented below apply to both universes unless otherwise stated.

NC Definition, All Procedures. Since the composition of a household is unchanged throughout its period of existence under NC, we have the following two possibilities:

- (a) No original sample people were in the household at any time during its period of existence, in which case the longitudinal household weight is zero for any time interval and procedure.
- (b) One or more original sample people were in the household throughout its existence, in which case the beginning and ending dates of the household are known, as is the composition of the household and complete data for each month of its existence. Consequently, features BW<sub>1</sub>, BD<sub>1</sub>, FW<sub>1</sub>, and FD<sub>1</sub> apply.

Furthermore, T<sub>1</sub> applies since procedures BH, BI, CM, and AW all reduce to the cross-sectional household weight at the beginning date of the household, while HW is the weight of the householder at the beginning date.

SH Definition, HW Procedure. The explanation is similar to the one given above, except

now the two cases are: (a) The householder was not an original sample person. (b) The householder was an original sample person.

SH and RM Definitions, BH Procedure. T<sub>1</sub> is applicable, since by definition the weight is the cross-sectional household weight as of the beginning date of the household. BW<sub>2</sub> applies because the longitudinal household weight is the cross-sectional household weight as of the first month in sample if the household began that month, while otherwise the weight will be zero since there were no original sample people in the household when it began. (For the restricted universe, households which entered sample after the beginning of the time interval always receive a zero weight.)

BD<sub>1</sub> holds since all households with positive weights were in sample at their beginning date and no retrospective subject-matter data is therefore needed.

FW<sub>1</sub> holds for the unrestricted universe since the weight is determined at the beginning date of the household. However, for the restricted universe, it is necessary to know if the household continued to exist throughout the entire time interval because it receives a zero weight for the time interval if it did not continue. Under current procedures a household which no longer has any original sample person is not followed, and it would therefore generally not be possible to determine if it remained in existence for the entire time interval. Consequently, FW<sub>2</sub> applies.

FD<sub>2</sub> applies since there would be missing data for all households with positive weights which continued to exist after there were no longer any original sample people present, which could happen for either of these definitions.

SH and RM Definitions, BI Procedure. T<sub>2</sub> is applicable since time intervals with different beginning dates may yield different longitudinal weights. BW<sub>1</sub> applies for the restricted universe, since the longitudinal weight is the cross-sectional household weight as of the first month of the time interval for all households in sample that month, and zero for all other households. However, BW<sub>2</sub> applies for the unrestricted universe since longitudinal households that entered sample after the beginning of the time interval are treated as in the BH procedure.

BD<sub>1</sub> holds since any household with a positive weight was either in sample the first month of the time interval or the month that the household began, and consequently, no retrospective data are needed.

As in the BH procedure, and for the same reasons, FW<sub>1</sub> applies for the unrestricted universe, FW<sub>2</sub> for the restricted universe and FD<sub>2</sub> for both universes.

SH and RM Definitions, CM Procedure, Restricted Universe. T<sub>2</sub> is applicable since any two intervals may yield different longitudinal weights.

Furthermore, BW<sub>1</sub>, BD<sub>1</sub>, FW<sub>1</sub>, and FD<sub>1</sub> apply. The explanation is similar to that given for the NC definition except now the two cases are: (a) No original sample people were household members for the entire time interval. (b) At

least one original sample person was a household member for the entire time interval.

SH and RM Definitions, AW Procedure. For an explanation for this row of the table see the original paper.

Utilization of Data. Having compared the procedures with respect to needs for unavailable data and the multiplicity of weights, we now turn our attention to variance. To compare the variance characteristics of the procedures we will focus on the amount of collected data that is used in obtaining estimates, since this is a primary determinant of variance. This discussion will also better illustrate the proportion of data needed for estimation that is unavailable for each procedure. In general, the greater this proportion is, the larger the burden is on any missing data procedure employed, with a resulting greater potential for bias problems. To make the comparison we show in Table 2, all 24 possible cases of how the data on a longitudinal household may be complete, partly available, or nonexistent for a particular time interval.

The symbols  $t_B$ ,  $t_1$ ,  $t_2$ , and  $t_E$  denote beginning date of household, first sample month, last sample month, and ending date of household respectively. The columns indicate different time intervals. Interval B is the interval of interest. Interval A is from  $t_B$  until the beginning of interval B, while interval C is from the end of interval B until  $t_E$ . The fifth case, for example, is of a household that formed before interval B about which we are missing some data pertinent to the early part of interval B. The first nine cases comprise the restricted universe. The last 15 cases fill out the unrestricted universe. Each case is marked as having complete data, partial data, or no data. Of course, all of this is assuming perfect response. The only type of missingness that we are discussing here is that caused by operational procedures. On the right there is a column for each procedure with an "A" entered if it always uses the case, an "S" if it sometimes uses the case but not always (which will be explained in the discussion that follows), and a blank otherwise. These comparisons do not apply to the NC definition, for which all five procedures use all the complete cases and no other cases.

Table 2.  
Data Utilization

Interval A	Interval B	Interval C	Completeness	Procedure				
				BH	BI	CM	AW	HW
1	$t_B=t_1$		perfect	A	A	S	A	S
2	$t_B<t_1$		perfect	A	S	A		
3	$t_B=t_1$	$t_2$	some missing	A	A		A	
4	$t_B<t_1$	$t_2$	some missing	A			A	
5	$t_B$	$t_1$	$t_2<t_E$				A	
6	$t_B$	$t_1$	$t_2$				A	
7	$t_B=t_1$	$t_2$	$t_E$				A	
8	$t_B<t_1$	$t_2$	$t_E$				A	
9	$t_B$		$t_1$					
10	$t_B=t_1$	$t_2=t_E$	perfect	A	A		A	S
11	$t_B<t_1$	$t_2<t_E$	perfect	A	A		A	
12	$t_B=t_1$	$t_2<t_E$	perfect	A	A		A	S
13	$t_B=t_1$	$t_2=t_E$	perfect	A	A		A	S
14	$t_B=t_1$	$t_2<t_E$	some missing	A	A		A	
15	$t_B<t_1$	$t_2<t_E$	some missing	A	A		A	
16	$t_B=t_1$	$t_2<t_E$	some missing	A	A		A	
17	$t_B=t_1$	$t_2$	$t_E$				A	
18	$t_B$	$t_1$	$t_2<t_E$				A	
19	$t_B<t_1$	$t_2<t_E$	some missing				A	
20	$t_B<t_1$	$t_2$	$t_E$				A	
21	$t_B<t_1$	$t_2<t_E$	some missing				A	
22	$t_B=t_1$	$t_2$	all missing	A				
23	$t_B<t_1$	$t_2$	all missing					
24	$t_B$	$t_1$	$t_2<t_E$					

The BH procedure uses the complete cases 1, 10, 12, and 13, but does not use the complete cases 2 and 11. It also uses the partial cases 3, 14, 16, and 17, and cases 7 and 22 for which there is no data in interval B. The BI procedure uses all the complete cases, more of the partial cases and none of the cases with no data. We thus think the BI procedure will tend to produce smaller variances than the BH procedure since it uses more of the available data. However, it is not clear in general which of these two procedures has the smaller proportion of needed data that is missing.

The CM procedure is appealing for the restricted universe since it uses all the complete cases (except in the rare situation when there is at least one original sample person present for every month of interval B, but none of them are present for the entire interval), and none of the other cases. It should thus have fairly small variances and has only the slight bias indicated in Section 3. However, it is not applicable to the unrestricted universe.

The HW procedure uses the same complete cases as the BH procedure, except it does not use these cases when the householder is not an original sample person, and it uses none of the other cases. However, it is not applicable to the RM, and most other longitudinal household definitions.

The AW procedure is the most aggressive in utilizing partial data. It uses all the complete and partial cases while avoiding the cases with no data. Also note that it assigns smaller weights, in general, to the partial cases than the complete cases. We believe it will tend to produce the smallest variances for most definitions, particularly in the unrestricted universe, but also tends to have the highest proportion of data that is needed for estimation but unavailable.

## REFERENCES

- Dicker, Marvin and Casady, Robert J. (1982), "A Reciprocal Rule Model for a Longitudinal Family Unit," American Statistical Association - Proceedings of the Social Statistics Association.
- McMillen, David B. and Herriot, Roger A. (1984), "Toward a Longitudinal Definition of Households," SIPP Working Paper Series, No. 402, Bureau of the Census, Washington, D.C.
- Nelson, D.D., McMillen, D.B., and Kasprzyk, D. (1984), "An Overview of the Survey of Income and Program Participation", SIPP Working Paper Series, No. 401, Bureau of the Census, Washington, D.C.
- Whitmore, R.W., Cox, B.G., and Folsom R.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.