The information contained in this appendix represents a draft report produced by Battelle on behalf of EPA's National Exposure Research Lab under Task Order 19 of Contract 68-D-99-011.

White Paper on Evaluation of Sampling Design Options for the National Children's Study

Appendix E

Methods to Deal with Nonresponse

The information contained in this appendix represents a draft report produced by Battelle on behalf of EPA's National Exposure Research Lab under Task Order 19 of Contract 68-D-99-011.

Disclaimer

The information contained in Appendix E was extracted from a draft report entitled "Development of Exposure Assessment Study Design for the National Children's Study: Draft Literature Search Report" prepared by Battelle, Harvard University, Westat, and the University of North Carolina as part of Task Order 19 under Battelle Contract 68-D-99-011 for the National Exposure Research Laboratory at the U.S. Environmental Protection Agency (EPA). This information reflects contributions to this draft report from Graham Kalton, Dwight Brock, G. Hussain Choudhry, Robert Clickner, and Adam Chu of Westat. This draft report has not yet been released by the EPA for public dissemination, and is currently under peer review prior to being released as an EPA report. Therefore, the information contained in this Appendix should not be cited or quoted.

The information contained in this appendix represents a draft report produced by Battelle on behalf of EPA's National Exposure Research Lab under Task Order 19 of Contract 68-D-99-011.

Compensating for Wave Nonresponse in Panel Surveys

E-1. INTRODUCTION

The NCS is planned to be a longitudinal study that will start with a sample of pregnant women of sufficient size to produce a sample of about 100,000 live births, who will then be followed for 21 years or more. Many waves of data will be collected on these births as they develop into adulthood. Inevitably in this process some of the data that should have been collected will be missing.

One source of missing data in any sample survey is noncoverage, which arises when some population elements are not included on the frame from which the sample is selected. Thus, some pregnant women may not appear on the sampling frame from which the NCS sample is drawn, and hence they and their children will not be represented in the NCS sample. Noncoverage is outside the scope of this discussion and will not be treated further here.

Other sources of missing data are various forms of nonresponse, of which three main types may be usefully distinguished for the NCS. First, some sampled pregnant women will refuse or for other reasons fail to participate in the enrollment interview and be lost to the study; in this case all their NCS data will be missing. Second, at any wave of data collection, there may be participation but some of the data items may be missing. And third, for some sample members none of the data may be collected for certain of the waves. All these categories except the last apply to all surveys, both cross-sectional and panel surveys. It is the last category—wave nonresponse—and the methods for handling it that are the focus of this review.

The main methods for handling wave nonresponse treated in this review are general purpose procedures that aim to compensate for the missing data in an overall way, so that a multitude of different analyses can be conducted with the resultant data file. This approach is the standard one for handling nonresponse in multi-purpose sample surveys. The alternative approach of incorporating nonresponse into a model for a specific analysis is briefly discussed at the end of this review.

The general purpose methods for handling wave nonresponse are essentially the same weighting adjustment and imputation methods that are used in cross-sectional surveys, but they are applied in a more complex form. It will therefore be useful to review these methods initially in order to provide a framework for the later discussion of wave nonresponse. Section 2 gives a brief review of survey nonresponse compensation methods to serve as this framework.

E-2. COMPENSATING FOR MISSING SURVEY DATA

Along the lines indicated above, survey nonresponse can be usefully classified as:

Total nonresponse: Eligible sample members who do not participate in the survey provide no survey data.

Item nonresponse: Some participating sample members fail to provide data for a small number of survey items.

Partial nonresponse: Some participating sample members fail to provide a large portion of the survey data.

One value of this classification is that it relates to the methods used to compensate for the missing data. As discussed below, weighting adjustments are generally used to compensate for total nonresponse and imputation is used for item nonresponse (i.e., inserting a value for the missing response). The distinction between total nonresponse and item nonresponse is widely made in the sampling literature. Partial nonresponse is less often classified separately, and may be viewed a set of item nonresponses. However, it is useful to make a separate class for partial nonresponse because the choice of weighting or imputation as the method of compensation is not straightforward for this class, and, whichever method is adopted, the form of the compensation is more complex.

Partial nonresponse commonly occurs when the survey data collection involves two or more components that require separate contacts with the sampled units. For example, the first wave of the Early Childhood Longitudinal Study—Birth Cohort (ECLS-B) has separate components for data on the child, on the parent, and on the father. Sometimes one or more of these components is missing. Also, of particular relevance here, wave nonresponse is a form of partial nonresponse, with data being collected for some but not all waves of a panel survey. Wave nonresponse gives rise to a sizable set of item nonresponses in the longitudinal file created by putting together the survey data for all waves of data collection. Like the ECLS-B, the NCS will face problems of partial nonresponse both because it will likely involve more than one data collection component at a given wave and because of wave nonresponse. This discussion is concerned with NCS wave nonresponse but some of the issues discussed apply also to missing components within a wave.

Consider first the estimation of the population mean $\overline{Y} = \Sigma^n Y_I / N$ of a variable *Y* in a population of size *N* in the case of complete response. In this case, the sample mean $\overline{y}_n = \Sigma^n w_i y_i / \Sigma^n w_i$ of a sample of size *n* is an approximately unbiased estimate of $\overline{Y}$, where w_i is the inverse of the selection probability of sampled element *i* (often termed the "base weight").

Now suppose that some sampled elements fail to respond. One approach is to model nonresponse as the outcome of an underlying random process where every individual in the population has a nonzero probability of response (see for example, Oh and Scheuren 1983). Under this formulation, the approximate bias of an estimated mean based on respondent data, $\overline{y}_r = \Sigma^r w_i y_i / \Sigma^r w_i$, where *r* denotes the respondents, can be expressed as

$$B(\overline{y}_r) = \frac{1}{N\overline{\phi}} \Sigma (Y_i - \overline{Y})(\phi_i - \overline{\phi})$$

where ϕ_i is and the response probability of the *i*th individual in the population, and $\bar{\phi}$ is the mean response probability in the population, i.e., the expected overall response rate (Brick and Kalton 1996). This relation shows that the nonresponse bias of the unadjusted respondent mean can be written as a covariance between the population characteristic of interest and the "propensity to respond" of the population members. Thus, the bias will be zero if the characteristic of interest, Y_i, is independent of response propensity, ϕ_i. In general, however, the bias of the respondent mean will not be zero. The purpose of nonresponse adjustments is to reduce this bias and ideally to eliminate it.

A common method for compensating for total nonresponse is to divide the overall sample into a set of classes (or cells), to compute the weighted response rate in each class, and to increase the base weights of respondents in a class by the inverse of the class response rate. Thus the adjusted weight of element *i* in class *c*, is $w'_{ci} = w_{ci}\phi_c^{-1}$, where $\phi_c = \Sigma^{r_c} w_{ci} / \Sigma^{n_c} w_{ci}$. The approximate bias of the respondent mean with the adjusted weights, $\bar{y}_{wc} = \Sigma^{r_c} w'_{ci} y_{ci} / \Sigma^{r_c} w'_{ci}$, is

$$B(\bar{y}_{wc}) = \frac{1}{N} \Sigma_c \Sigma_i \bar{\phi}_c^{-1} (Y_{ci} - \bar{Y}_c)(\phi_{ci} - \bar{\phi}_c) \; .$$

This bias is zero if either (a) the Y_{ci}'s are constant, or (b) the ϕ_{ci}'s are constant within each class (or, more generally, if the Y_{ci}'s and the ϕ_{ci}'s are uncorrelated within the classes). A goal of nonresponse adjustment procedures, therefore, is to identify classes of the population within which either of these two conditions is satisfied. If the focus is on condition (a), such weighting classes will be specific to a particular survey variable, and different classes would need to be sought for each variable. As a result, different sets of weights would be needed for each variable. However, it is impracticable to have several sets of weights for general purpose analyses involving the relationships between variables. Thus, the usual approach for identifying relevant weighting cells for nonresponse adjustment is one of finding classes that satisfy the second condition.

The information contained in this appendix represents a draft report produced by Battelle on behalf of EPA's National Exposure Research Lab under Task Order 19 of Contract 68-D-99-011.

In most population surveys little information is available for the total nonrespondents. Often the only information is the primary sampling unit and the stratum in which the sampled household is located. There is therefore limited information for use in forming weighting classes. In such cases the classes can be simply formed as the cells of a cross-tabulation of the available variables. When there is more information available, it may be necessary to conduct analyses to choose the most effective variables to use in the nonresponse weighting adjustments and also to use alternative forms of adjustment than the weighting class adjustments described above. The situation in which a large amount of auxiliary information is available arises in particular with partial nonresponse, and will be taken up later in relation to weighting adjustments for wave nonresponse in panel surveys.

Dropping total nonrespondents from the analytic data file and compensating for them by a weighting adjustment involves no real loss of data. They have not responded to any of the survey variables and the weighting classes capture most of what is known about them. The situation with item nonresponse is different. Dropping cases with item nonresponse from the data file would involve a major loss of data, since the responses to the many items these respondents have answered would be lost. For this reason, imputation rather than weighting is used to compensate for item nonresponses. Imputation is a process of assigning values for missing item responses. The imputed values are derived using the responses to other items available for the respondent (and any other auxiliary information that is relevant) as predictors of the missing values. Numerous methods have been developed for imputation, with a wide range of properties. However, as noted by Kalton and Kasprzyk (1986), most of these methods fall within a general multiple regression framework.

Let y_{mi} be a missing value of characteristic *y* for sample individual *i*, and suppose auxiliary information $\mathbf{z_i}$ $z_{i1},\ldots,z_{ip}$ is available for that individual (i.e., frame information and responses to other survey items). Many imputation methods (including hot deck procedures, class mean imputation, regression imputation) can be represented by the following equation:

$$\hat{y}_{mi} = b_{r0} + \sum_{j=1}^{p} b_{rj} z_{mij} + \hat{e}_{mi} \quad .$$

The b_{r0} through b_{rp} values are estimated regression coefficients derived from the observed data. $\hat{e}_{mi}$ is a residual term which varies by method. For deterministic imputation methods (mean imputation, nonstochastic regression imputation using a predicted value as the imputed value), this term is zero. For stochastic imputation methods, this term is nonzero, with its value depending on the actual imputation method. Since deterministic imputation methods distort the distribution of the variable subject to imputation, stochastic methods are generally preferred.

A major concern with imputation is that its use can distort the associations between the variable being imputed and other variables in the data set. If variable *y* has some imputed values and in the analysis its association with *x* is estimated, the *xy* association will be attenuated towards zero unless *x* is used as an auxiliary variable in the imputation method. Given the large number of variables in a typical cross-sectional survey, maintaining all the associations of potential interest to analysts is a major challenge for imputation. That challenge becomes much greater still with a panel survey since analysts are interested in associations of variables both within and across waves of the survey. Another aspect of the challenge of imputation is handling missing responses for a range of different variables, and with different patterns of missingness for different respondents (sometimes known as the "Swiss cheese problem"). This is the usual situation encountered in practice and it adds to the challenge of carrying out the imputations in a way that does not distort associations. It is concerns about distorting associations that lead many analysts to prefer weighting adjustments to imputation for handling most forms of wave nonresponse in panel surveys. They prefer to jettison data for respondents who fail to participate in all waves, and use weighting adjustments for such cases, rather than risk the distortion of associations that can arise with mass imputation of all the items in the missing wave.

The information contained in this appendix represents a draft report produced by Battelle on behalf of EPA's National Exposure Research Lab under Task Order 19 of Contract 68-D-99-011.

E-3. WAVE NONRESPONSE IN PANEL SURVEYS

Wave nonresponse is a type of partial nonresponse that arises often in panel surveys. It is common for sample units that cooperate in a panel survey to fail to provide data for one or more of the survey waves. Some sample members may drop out of the survey at one wave and be lost for all subsequent waves (attritors), while others may miss one wave but return to the panel at some subsequent wave (nonattritors). Table A summarizes some typical patterns of wave response (X) and nonresponse (0) for a panel survey with five waves of data collection. The attrition nonrespondents exemplified by patterns 2 through 5 of the table are "terminal" nonrespondents who do not re-enter the study after the first wave in which they are nonrespondents. The nonattrition nonrespondents, on the other hand, respond in some waves, with a gap (or more than one gap) of one or more waves of nonresponse. Note that individuals who become ineligible for the survey in a particular wave are not considered to be nonrespondents even though they provide no survey data for that wave.

Table E-A. Illustrative reporting patterns for a panel survey with five data collection waves

		Wave				
Pattern	Response status	1	2	3	4	5
1	Respondents	X	X	X	X	X
2	Attrition nonrespondents	X	X	X	X	0
3		X	X	X	0	0
4		X	X	0	0	0
5		X	0	0	0	0
6	Nonattrition wave nonrespondents	X	X	0	X	X
7		X	0	0	X	X
8		X	0	0	0	X

An X indicates response and an 0 indicates nonresponse.

Not represented in Table A are the total nonrespondents who provide data for none of the panel waves. In many panel surveys, nonrespondents in the first wave of data collection are not followed up in future waves. Often they are not followed because they have not been clearly identified and because no data are available for tracing them. In this situation, all first wave nonrespondents are total nonrespondents. This is, for example, the case for the U.S. Census Bureau's Survey of Income and Program Participation (SIPP), the Medicare Current Beneficiary Survey (MCBS) conducted by the Centers for Medicare and Medicaid Services (CMS), and the Early Childhood Longitudinal Study – Birth Cohort (ECLS-B).

Most panel surveys set rules to specify when attempts to follow wave nonrespondents should cease. For example, the following rules for the MCBS, which employs a rotating panel design that includes 12 waves of data collection (e.g., see O'Connell, Chu, and Bailey, 1997), specify that wave nonrespondents at one wave are followed in the next wave, but if they are nonrespondents on the next wave also, they are not followed in later waves. Thus wave nonresponse is limited to only one consecutive occurrence before the beneficiary is dropped from the study fieldwork entirely and treated as a nonrespondent on later waves. Similarly, in SIPP households that miss two consecutive waves are not followed in subsequent waves. Thus, pattern 6 in Table A would be permissible for SIPP and MCBS, but not patterns 7 or 8.

Table A also points out the types of respondents that would be available for various types of analysis (assuming that imputation is not used for the missing waves). Consider first cross-sectional analyses of individual waves. Analysis of wave 1, for example, could include all individuals with reporting patterns 1 through 8, with weights adjusted for the loss of the initial nonrespondents. Weighting adjustment procedures outlined in the previous section would be appropriate in this case, with the auxiliary data used to define adjustment classes being limited to design information available for both respondents and nonrespondents. Similarly, analysis of wave 2 in the example of Table A would involve individuals with reporting patterns 1-4 and 6, while individuals with patterns 5, 7, and 8 would be treated as nonrespondents. In this case, either weighting adjustments or imputation could be used to compensate for the wave 2 losses, but unlike the adjustment for wave 1, there is a wealth of relevant survey data from wave 1 to employ in the adjustment procedures. Provided that the wave 1 nonrespondents are not followed in subsequent waves, wave 1 data will be available for all wave 2 losses. As a final example, consider analysis of wave 5. Here, the respondent sample would include individuals with reporting patterns 1, 6, 7, and 8. However, adjustment for nonresponse will be complicated by the fact that the nonrespondents have varying amounts of data from prior waves.

Now consider longitudinal analyses of the survey data. An analysis of data from both waves 1 and 5 could be conducted using respondents in patterns 1, 6, 7, and 8, using data from wave 1 in the compensation procedure for the other patterns. An analysis of data from waves 2 and 5 could use data from patterns 1 and 6. The adjustment procedure in this case could use data from wave 1 for all the other patterns, and also wave 2 data for patterns 2, 3, and 4. Finally, consider a longitudinal analysis that involves data from all five waves. Only pattern 1 has all the data available for such an analysis, and adjustments would be needed for all other patterns. It should, however, be noted that some forms of longitudinal analyses can tolerate some missing data (e.g., growth curve analyses). Thus, some patterns in addition to pattern 1 may be included for such analyses.

As this discussion illustrates, competing analytic needs can lead to many different sets of weights that depend on the particular patterns of nonresponse. With *t* waves, there are potentially $2^t - 1$ combinations of waves that could be of analytic interest, and this number grows rapidly with *t*. There are obvious cost implications associated with developing separate weights for every possible analytic subset. Moreover, using different sets of weights in analysis can be operationally awkward. unwieldy, and error prone. An alternative solution is therefore needed.

One solution is to convert nonattrition nonreponse to attrition nonreponse by ignoring data collected at waves later than the first nonresponding wave (e.g., ignoring the data for waves 4 and 5 in patterns 6 and 7 and for wave 5 in pattern 8). This has two advantages. First, it reduces the potential number of sets of weight from $2^t - 1$ to t. Second, attrition nonresponse is much easier to handle with weighting adjustments than nonattrition nonresponse. Weighting adjustments for attrition nonresponse are discussed in Section 4.

A disadvantage of the above solution is that it can discard a large amount of data when the panel has a sizable number of waves (e.g., data for both waves 4 and 5 with patterns 6 and 7).

An alternative solution is to use imputation to fill in the values for missing waves, or at least for certain patterns of missing waves. As noted earlier, with imputation, there is a serious problem of possibly distorting associations between variables, but that may be preferable to discarding large amounts of data. Imputation for missing waves is the subject of Section 5.

E-4. COMPENSATING FOR ATTRITION

Adjustment for attrition nonresponse is generally more straightforward because the nonresponse is "monotone." Each successive wave adds an additional set of nonrespondents on top of those that dropped out in previous waves. Weighting adjustments can be made using a wave-specific nonresponse adjustment that is applied to the nonresponse adjusted weights from the previous wave. An advantage of this approach is that the nonresponse adjustment for the current wave can be developed independently of the adjustments developed in previous waves. Thus, the overall weight for wave *t* is a product of the form:

$$w_{it} = w_i \hat{R}_{i1}^{-1} *** \hat{R}_{it}^{-1}$$

where w_i is the base weight for sampled unit *i* and $\hat{R}_{it}$ is the weighted response rate for the weighting class to which unit i is allocated at wave *t*. Note that for $t \geq 2$, $\hat{R}_{it}$ is the "conditional" response rate among those individuals who responded in wave *t*–1. Each $\hat{R}_{it}$ can be computed as follows within a set of weighting classes that are appropriate for attrition in that period:

$$\hat{R}_{it} = \frac{\sum_{j \in c_i(t)} I_{jt} w_{j(t-1)}}{\sum_{j \in c_i(t)} w_{j(t-1)}}$$

where $c_i(t)$ denotes the adjustment cell containing sample member *i* in period *t*, I_{jt} is a dichotomous indicator for response in period *t* (i.e., is 1 if sample unit *j* continues to respond in period *t*, and is 0 otherwise), and $w_j(t-1)$ is the weight for sample unit *j* adjusted for attrition up to period *t*-1. Note that the $\hat{R}_{it}$ factors will be very close to 1 if wave-to-wave attrition is small.

To develop adjustments that are effective in reducing bias it is important that the weighting classes be defined in a way that attempts to ensure that each individual in the class has the same response propensity. For the first-period (total) nonrespondents, relevant information about the nonrespondents may be limited to information about the sample units available on the frame. For later-period attritors information from the previous waves will provide a wealth of information that can be used to inform the adjustment process. In this case there will likely be so much information that an initial task will be to select the most relevant variables from a large set of potentially useful ones.

There are a number of general statistical methods that can be used to determine the required adjustment classes. These include logistic regression methods and tree-based methods such as CHAID (Magidson 1993) and CART (Breiman et al. 1993). In the logistic regression framework, the propensity of response, ϕ_{it}, is assumed to have the form:

$$\log\left(\frac{\phi_{it}}{1-\phi_{it}}\right) = x_{it}\beta_t$$

where x_{it} is a vector of characteristics for sample unit *i*, and β_t is the corresponding vector of coefficients (parameters). Estimates of β_t and ϕ_{it} are generated using weighted logistic regression where I_{it} is the dependent variable, x_{it} is a vector of predictor variables, and $w_{j(t-1)}$ is the weight in the logistic regression.

As stated above, in selecting models for later period attritors, the set of potential predictor variables will tend to be very large, as it may include all items collected in previous waves. Fitting the model may therefore require a stepwise approach (see for example Rizzo, Kalton, and

Brick, 1996) to reduce the initial set of independent variables to a parsimonious set of predictors. Rizzo et al. (1996) also discuss methods for forming the final cells from the logistic regression models by pooling together sample units with similar predicted ϕ_{it} values. Alternatively, it is possible to forgo the use of weighting cells entirely and use the reciprocal of the predicted ϕ_{it} from the model as the nonresponse adjustment, but this could lead to very large weighting adjustments for sample units with very small ϕ_{it} values. Moreover, such adjustments are not robust to misspecification of the model.

The tree-based methods (CHAID, CART, or others) can either be used in conjunction with logistic regression, or alone, to select a set of weighting adjustment cells. These methods allow one to find important interactions, and develop cells which reflect these interactions. For example, if females and males differ in response rate among African Americans, but not in the overall population, then logistic regression couldleave out sex as a cell-determining characteristic. However, the tree-based methods will subset African Americans by sex. A drawback of the tree-based algorithms is that they tend not to be robust (i.e., small changes in the data can lead to large changes in the selected tree).

Alternatives to weighting-cell adjustment as described above are raking and GREG (see, for example, Kalton and Flores-Cervantes 2003). The weighting-cell adjustment method can be thought of in the following way. Initial weights for each sample unit in period *t* are given by $w_{j(t-1)}$ for respondents and nonrespondents in period *t*. After sample losses due to attrition, we want to adjust the weights of the respondents in period *t* to represent themselves and the period *t* attritors, setting the weights of the attritors to 0. The idea is to determine a new set of weights, w_{it}, within each weighting cell such that when summed over the respondents equals the sum of the original weights when summed over the entire sample. In weighting-cell adjustment, this is done by multiplying each respondent's $w_{j(t-1)}$ weight by the ratio .

In raking and GREG, the respondents' weights in period *t* are adjusted to match sample unit summations of the $w_{j(t-1)}$, but not for a simple set of cells. The w_{it} weights are adjusted to be "as close as possible" to the $w_{j(t-1)}$ weights, in the sense of a well-defined distance measure, while still respecting a set of constraints (see, for example, Deville and Sarndal 1992 and Deville, Sarndal, and Sautory, 1993). In raking, these constraints are marginal cell totals for two or more dimensions (the marginal cell totals are summations of the $w_{j(t-1)}$ weights over all sample units in each marginal cell). The weights are iteratively adjusted for each set of margins until they simultaneously satisfy the target summations for each cell in each dimension. The original reference for raking is Deming and Stephan (1940).

GREG is similar except that the constraints are not simple summations of weights necessarily but are instead summations of weighted sums of auxiliary variables (which are known for both respondents and nonrespondents). The new weights $w_j(t)$ are adjusted using regression predictors so that the weighted sums using the new weights are equal to the weighted sums of the $w_i(t-1)$ weights over the full set of respondents and attritors. See for example Fuller (2002).

Folsom and Witt (1994) also discuss alternative nonresponse adjustments based on a generalized raking method (adjusting weights for responding panel members to totals computed for all sample members). Their raking method is related to the Logit (L,U) method of Deville, Sarndal, and Sautory (1993), which smoothes large calibration adjustments. An et al. (1994) examine another alternative which simultaneously adjust the weights to two sets of constraints: the sample totals and external control totals. This adjustment is a "full information" procedure which utilizes the two sets of constraints in a more efficient manner.

The information contained in this appendix represents a draft report produced by Battelle on behalf of EPA's National Exposure Research Lab under Task Order 19 of Contract 68-D-99-011.

E-5. COMPENSATING FOR NONATTRITION NONRESPONSE

The potential patterns of nonattrition nonresponse are often restricted by a panel survey's followup rules. Indeed, if the rules specified that wave nonrespondents should not be followed after the first nonresponding wave, there would be no nonattrition nonresponse. Such a rule is sometimes adopted. However, in general there will be cases of nonattrition nonresponse.

As noted earlier, nonattrition nonresponse can be handled by weighting adjustments, either by developing weights for the set of respondents who provide data for all waves involved in the analyses being conducted, or by converting nonattrition nonresponse to attrition nonreponse. The first of these methods may be the preferred solution when key analyses involve specific waves. For example, if a key objective is to relate outcomes for the sampled youth at age 20 to events in pregnancy, nonresponse for the intervening years is not an issue. Weighting adjustments can be generated for all sampled youth with complete data in the relevant waves (and with the relevant items), thus making maximal use of the data collected relating to those analyses. However, the proliferation of weights with this solution is highly problematic for general purpose analyses of combinations of waves. Converting nonattrition nonresponse to attrition nonresponse jettisons a good deal of data for cases with only one or two missing waves interspersed with responding waves. Given the disadvantages associated with either of the weighting solutions, imputation is sometimes used for certain kinds of missing waves in nonattrition cases.

For example, the SIPP uses longitudinal imputation to assign values for whole waves of missing data in its longitudinal data files that contain data for all waves. Longitudinal imputation was used initially to impute for one missing wave provided data were available for the preceding wave and the succeeding wave. This imputation method, when applied to the 1991, 1992, and 1993 Panels, resulted in the retention of an additional 5-8 percent of the sample persons in the panel files. Beginning with the 1996 Panel, SIPP expanded the use of longitudinal imputation to include persons with two consecutive missing waves bounded by responding waves. SIPP collects data at 4-monthly intervals, and at each wave it collects data for each of the four preceding months separately. The imputation procedure used in the SIPP is essentially as follows. When the last month of the preceding wave and the first month of the succeeding wave have the same value for an item, then that value is imputed for every month in the missing wave. If the last month of the preceding wave and the first month of the succeeding wave have different values for an item, then a randomized procedure is used to select a "change" month. One of the months of the missing wave period is selected with equal probability for each household as the change month for the household. For all months in the missing wave up to the randomly selected change month, the value from the last month of the preceding wave is imputed. For all months from the change month onwards, the value from the first month of the succeeding wave is imputed. The SIPP Quality Profile (U.S. Bureau of the Census1998) describes the extensive research into longitudinal imputation that led to these imputation methodologies.

A similar approach is used in the MCBS to compensate for a missing wave. In the MCBS, annual cost and payment statistics are produced for each calendar year. Each calendar year comprises three waves of data collection. Thus, if a respondent did not report cost and payment data for all three waves, the missing data were imputed using item-specific averages derived from available reported data. This general approach was used for all relevant items in the missing wave(s). Additional information about the MCBS imputation procedures are given in Centers for Medicare and Medicaid Services (2001).

If imputation is chosen to fill in particular missing waves, then there are a variety of methods available (see, for example, Kalton 1986 and Lepkowski 1989). As discussed in Lepkowski (1989), the best approach for partial nonresponse may often be a judicious mix of imputation and weighting adjustment, based on the needs of the users and the resources available. Imputation can be used to fill out particular missing waves when the number of these waves is limited, with weighting adjustment used to adjust for other strings with a larger number of missing periods.

The information contained in this appendix represents a draft report produced by Battelle on behalf of EPA's National Exposure Research Lab under Task Order 19 of Contract 68-D-99-011.

E-6. ADDITIONAL EXAMPLES OF NONRESPONSE ADJUSTMENT IN LONGITUDINAL SURVEYS

E-6.1 NATIONAL EDUCATIONAL LONGITUDINAL SURVEY (NELS - 88)

The original wave of the National Center for Education Statistics study, the National Educational Longitudinal Survey (NELS), was a nationally-representative, multi-stage probability sample of 8th grade students selected in 1988 with follow-up surveys conducted every two years thereafter. The base year design oversampled private schools and schools with higher than average enrollments of Hispanic and Asian students. As with most studies of its type, the NELS-88 experienced a considerable amount of attrition between waves, especially between the baseline and first follow-up (i.e., between 8th and 10th grades). Most of the sample attrition between waves 1 and 2 was due to budget constraints placed on the sponsoring agency. Therefore, it was believed that, since the data were missing by design, the effect of the missing data was not as closely related to the quantities being measured as in other studies and the nonresponse bias was small. Nevertheless, in several published analyses of this survey (see, for example, Lee and Smith 1995, Kao and Tienda 1998, and Rojewski and Yang 1997), longitudinal weights for each wave were calculated to inflate the sample values up to the population level. In addition, the weights provided an adjustment for wave nonresponse at each wave of the study.

In other studies of a similar type researchers have used modeling as well as weighting to account for missing data. For example, Baltagi (1998), following work by Verbeek and Nijman (1992), has suggested the use of three simple auxiliary variables in random effects regression models used to analyze longitudinal data:

> The number of waves in which the ith individual *participates* in the panel;
> A binary variable that takes the value 1 if the ith individual is observed over the entire length of follow-up of the survey; and

A binary variable indicating whether the ith individual responded in the last wave of the survey.

If these variables make a significant contribution to the underlying regression equation, then the effect of selection bias cannot be ignored in the analysis. Examples of studies in which Baltagi suggested the use of this technique included the Panel Study on Income Dynamics and the National Longitudinal Study.

E-6.2 NATIONAL LONGITUDINAL SURVEYS OF YOUTH

The National Longitudinal Survey of Youth 1979 (NLSY79) is a nationally representative sample of 14-22 year olds first surveyed in 1979. This is a part of the National Longitudinal Surveys program funded by the Bureau of Labor Statistics. As the sample members of the Original Cohorts (begun in 1966) aged, and as new federal legislation expanded employment and training opportunities for youth, this study was started to provide data to replicate earlier studies of labor market experiences of Americans. The surveys also contain a wide range of questions covering educational attainment, training investments, employment history, income and assets, welfare receipts, child-care costs, insurance coverage, health conditions, workplace injuries, alcohol and substance abuse, sexual activity, and marital and fertility histories. The NLSY79 was administered annually, starting in 1979, then biennially since 1994.

Davey, Shanahan and Schafer (2001) describe an analysis of this survey in which they corrected for selective nonresponse using multiple imputation techniques. First they fitted logistic regression models to predict missing/observed status on the patterns of missing data in the study. Predictors of missing data included mother's age at the time of birth, past duration of time the family was in poverty, number of transitions into poverty, mother's marital status, mother's age at the time of the follow-up, and children's psychosocial adjustment. Recognizing that missing data were likely to influence the outcome of the analysis, the authors then proceeded to compare list-wise deletion of cases with missing data to multiple imputation for adjusting for the nonresponse, which allowed retention of individuals with partially missing data in the analysis.

E-6.3 ESTABLISHED POPULATIONS FOR THE EPIDEMIOLOGIC STUDY OF THE ELDERLY (EPESE)

These studies were sponsored by the National Institute on Aging of the National Institutes of Health. The purpose was to study changes in demographic and health characteristics of persons aged 65 and older in four communities based on annual interviews conducted over the period 1981-88. One particular emphasis of the EPESE was to track changes in functional health and disability and risk factors for these changes. During the course of the study a number of participants dropped out of the study at one or more waves but returned at a later wave. To account for these kinds of gaps in the data, Beckett Brock, Scherr and Mendes de Leon (1993) and Beckett, Brock, Lemke, et al. (1996) applied a Markov transition model to data from these four studies, two of which were based on complex sample surveys. The objective was to study the effects of several independent predictor variables on changes over time in the participants' ability to perform any one of four activities: transferring from a bed to a chair; walking across a small room; climbing a flight of stairs; and walking half a mile.

The use of the Markov model with an imbedded logistic regression allowed the researchers to calculate the likelihood of transitions through all possible paths the participants could have taken across the gaps in the data created when the participants dropped out of the study temporarily. The Markov model was also adapted to account for the complex designs used in two of the study sites by estimating the standard errors of the model parameters using a Taylor series approximation. An evaluation of the technique showed that ignoring the gaps or the complex design would have resulted in biased estimates of the model parameters and their standard errors.

The information contained in this appendix represents a draft report produced by Battelle on behalf of EPA's National Exposure Research Lab under Task Order 19 of Contract 68-D-99-011.

REFERENCES

An, A., Breidt, F., and Fuller, W. (1994). Regression weighting methods for SIPP data. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 434-439.

Baltagi, B. H. (1998). Panel data methods. In *Handbook of Applied Economic Statistics.* A. Ullah and D.G.A. Giles (eds.), pp. 291-323. New York: Marcel Dekker.

Beckett, L. A., Brock, D. B., Scherr P. A. and Mendes de Leon, C.F. (1993). Markov models for longitudinal data from complex samples. *Proceedings of the American Statistical Association, Section on Survey Research Methods,* pp. 921-925.

Beckett, L. A., Brock, D. B., Lemke, J. H., Mendes de Leon, C. F., Guralnik, J. M., Fillenbaum, G. G., Branch L. G., Wetle, T. T., Evans, D. A. (1996). Analysis of change in self-reported physical function among older persons in four population studies. *American Journal of Epidemiology,* vol. 143(8), pp. 766-778.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1993). *Classification and Regression Trees*. New York: Chapman & Hall.

Brick, J. M., and Kalton, G. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research 5,* 215-238.

Centers for Medicare and Medicaid Services (2001). CY 1999 Cost and Use Public Use File Documentation, November 2001, Baltimore, MD.

Davey, A., Shanahan, M. J., Schafer, J. L. (2001). Correcting for selective nonresponse in the National Longitudinal Survey of Youth using multiple imputation. *Journal of Human Resources,* vol. 36 (3), pp. 500-519.

Deming, W. E., and Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics, 11,* 427-444.

Deville, J-C., and Sarndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association, 87*, 376-382.

Deville, J-C., Sarndal, C.-E., and Sautory, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association, 88,* 1013-1020.

Fay, R. (1991). A design-based perspective on missing data variance. *Proceedings of the 1991 Annual Research Conference, U.S. Bureau of the Census*, 429-448.

Folsom, R., and Witt, M. (1994). Testing a new attrition nonresponse adjustment method for SIPP. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 428-433.

Fuller, W. (2002). Regression estimation for survey samples. *Survey Methodology 28*, 5-23.

Kalton, G. (1986). Handling wave nonresponse in panel surveys. *Journal of Official Statistics, 2,* 303-314.

Kalton, G., and Flores-Cervantes, I. (2003). Weighting methods. *Journal of Official Statistics,* forthcoming.

Kalton, G., and Kasprzyk, D. (1986). The treatment of missing survey data. *Survey Methodology 12*, 1-16.

Kao, G. and Tienda, M. (1998). Educational aspirations of minority youth. *American Journal of Education*, vol. 106 (3), pp. 349-384.

Lee, V. E. and Smith, J. B. (1995). Effects of high school restructuring and size on early gains in achievement and engagement. *Sociology of Education,* vol. 68 (4), pp. 241-270.

Lepkowski, J. (1989). Treatment of wave nonresponse in panel surveys, in *Panel Surveys* (D. Kasprzyk, G. Duncan, G. Kalton, and M. P. Singh, eds). New York: John Wiley & Sons.

Little, R. J. A., and David, M. H. (1983). Weighting adjustments for non-response in panel surveys. Working paper, U.S. Bureau of the Census, Washington, DC.

Little, R. J. A., and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley and Sons.

Magidson, J. (1993). *SPSS for Windows CHAID Release 6.0*. Belmont, MA: Statistical Innovations Inc.

Oh, H. L., and Scheuren, F. S. (1983). Weighting adjustments for unit nonresponse, in *Incomplete Data in Sample Surveys, Vol II: Theory and Annotated Bibliography* (W. G. Madow, I. Olkin, and D. B. Rubin, eds.). New York: Academic Press.

Rizzo, L., Kalton, G., and Brick, J. M. (1996). A comparison of some weighting adjustment methods for panel nonresponse. *Survey Methodology 22(1)*, 44-53.

Rojewski, J. W. and Yang, B. (1997). Longitudinal analysis of select influences on adolescents' occupational aspirations. *Journal of Vocational Behavior,* vol. 51, pp. 375-410.

U.S. Bureau of the Census (1998). *SIPP Quality Profile 1998*. 3rd ed. Washington, D.C.: U.S. Bureau of the Census.

Verbeek, M. and Nijman, T. E. (1992). Testing for selectivity bias in panel data models. *International Economic Review,* vol. 33, pp. 681-703.