NON-RESPONSE AND IMPUTATION

R. Platek and G.B. Gray[1]

The problems of dealing with non-response at various stages of survey planning are discussed with implications for the mean square error, practicality and possible advantages and disadvantages. Conceptual issues of editing and imputation are also considered with regard to complexity and levels of imputation. The methods of imputation include weighting, duplication, and substitution of historical records. The paper includes some methodology on the bias and variance.

1. INTRODUCTION

The reliability of survey estimates is governed by many factors, one of which is the effect of missing and inconsistent or incomplete data. Any survey, whatever its nature, suffers from some non-response or responses which fail data edit procedures. The question that should be answered is "what should we do with this kind of incompleteness in the data"? One can argue, of course, that if the magnitude of deficient data is less than one percent, one might not worry about it at all. But in practice, the size of non-response is more like 10%, 15% or more, depending on the subject matter.

To disregard the effect of non-response of such size may lead to survey results of unacceptable quality and it will definitely mean that population totals could not be estimated since they would be based on partial data only. On the other hand, the reliability of averages and proportions will be affected less than that of totals by non-response and one can also argue, with some justification, that in general, the effect of non-response on national estimates will be smaller than for some sub-national levels. Nevertheless, the elimination and the reduction of the effect of non-response and invalid responses is very

[1] R. Platek, and G.B. Gray, Household Surveys Development Division, Statistics Canada.

important and it should be undertaken at various stages of survey design as well as in the field. Despite these efforts, however, some non-response and deficiencies will remain in the data and, in practically all surveys, some form of adjustment or imputation for non-response will have to be considered.

Imputation may be defined as the assignment of data to empty fields (including total non-response) or a replacement of invalid data following certain rules. There is no known unbiased method of imputing unless several assumptions are made regarding non-respondents and respondents. There is, however, some evidence that certain methods may be more efficient than others.

2. DEALING WITH NON-RESPONSE

(i) Survey Planning and Development

At the planning stage, an awareness of the effect of non-response on the Mean Square Error of survey data will undoubtedly lead to a survey design with as little non-response as possible. Consequently, one of the important factors in planning a survey is a decision on the tolerance level of non-response and an experienced survey designer can estimate fairly accurately the level of response for a particular survey that can be expected under various survey conditions. It can be argued that for some surveys when only national estimates are required and when the characteristics of non-respondents are not strikingly different from those of respondents, a non-response rate (20-30%) may be tolerated even though this will result in an increase in sampling and perhaps in response variance. The same arguments can be applied to surveys whose objective is to provide some notion about trends and proportions. However, for surveys whose estimates must be precise and are required at various sub-national levels, the non-response rate should be kept as low as 5% or less and pockets of large non-response in local areas should also be avoided.

The survey cost is another item which will affect many factors in survey development including non-response. It is important to balance the other factors against the cost so as to achieve a non-response rate sufficiently low to serve the goals of the survey. It should also be realized that within reasonable limits, it is sometimes better to accept a somewhat smaller sample than originally planned and to transfer the resources to appropriate data collection, follow-up and estimation procedures. This would be particularly advantageous if the survey designer suspects large differences between respondents and non-respondents in their characteristics.

Apart from intuition and experience which undoubtedly play an important part in survey planning and development, one can identify a number of factors which are important in the design of surveys. These factors can be classified into three groups:

Group I
- a) sample size
- b) stratification
- c) degree of clustering
- d) sample allocation
- e) method of selection

Group II
- a) sampling frame
- b) method of interviewing
- c) selection, training and control of staff
- d) length of questionnaire and wording
- e) sensitivity of questions
- f) type of area in which the survey is taken
- g) feasibility of call-backs and the number of them
- h) publicity

Group III
- a) edit and imputation
- b) estimation
- c) variance estimation and other data analysis.

All these operations certainly affect the Mean Square Error to a varying degree. It is true that in practice we often lack enough data on the effect of most of the factors. However, since not all these factors are of equal importance, an examination of the more important components of the Mean Square Error would be very helpful. Let us suppose that the Mean Square Error can be decomposed into the following components:

$$MSE = V_S + V_R + V_{CR} + (B_S + B_R)^2$$

where

V_S = sampling variance
V_R = response variance
V_{CR} = correlated response variance
B_S = sampling bias
B_R = response bias.

Sampling variance (V_S) and sampling bias (B_S) are affected by all the factors in Group 1, by estimation procedures and also by the size of non-response. The larger the size of non-response, the greater the effect it has on sampling variance and bias. For example, since the sampling variance of the estimates is inversely proportional to the response rate in the case of a simple random sample, estimates based on such a sample with 80% response rate will have a sampling variance that is 12.5% higher than the variance of corresponding estimates with 90% response rate. In multi-stage clustered samples, the same relationship holds approximately but it affects mainly the final stage of sampling. The relationship between the bias and the size of non-response, while perhaps more important, is less obvious since it depends on both the magnitude of non-response and the characteristics of both respondents and non-respondents. In considering non-response it has to be taken in account that a reduction of non-response in the field does not necessarily ensure a reduction in bias. In fact, if the procedures for the reduction of non-response are not well thought out and appropriately executed, the bias may not be reduced and could even be increased.

In some surveys, survey conditions may affect the sampling variance and sampling bias. For example, the wording of the questionnaire and/or the training of the interviewers may operate in such a manner that legitimate extreme values are eliminated. A low sampling variance but a high sampling bias may result. The artificially low sampling variance may occur because the variance between units of the expected responses without extreme values will be lower than the variance between the true values with the extreme values. The extreme values on opposite sides of the mean value will not necessarily balance so that a high sampling bias could result. Consequently, the survey conditions may affect sampling variance and sampling bias.

Non-sampling components of Mean Square Error (V_R, V_{CR}, B_R^2) which also include non-response are affected to a varying degree by all the factors in Group II. In addition, the Mean Square Error is also affected by some factors in Group I. For example, clustering may affect the correlated variance in much the same way as it affects sampling variance since households in clusters may produce higher correlations in response errors than households further apart. Since the estimate is a function of the observed values, which in turn are subject to non-sampling error, and since each distinct estimation procedure involves a different function, then the non-sampling variance will also be affected by the estimation procedure.

(ii) Data Collection Stage

Non-response can be reduced by persistent efforts of interviewers and by motivation of non-respondents to become respondents. The persistent efforts are usually in the form of repeated attempts to contact a respondent and to gather information about him or her. There is a point beyond which it is impossible to attempt further callbacks, either because the survey is to be completed by a certain date or because there are not sufficient

funds. In the case of telephone interviewing, the cost is only that of repeated telephone attempts and with mail surveys that of subsequent reminders. However, in the case of personal interviews where, for reasons of cost, the sample must usually be clustered to minimize travelling time and distance between successive calls, repeated callbacks often result in a greater distance between households and the cost per unit may become unacceptably high, without any reductions in the variance.

Further, if the probability of non-response were the same for each unit, the non-respondents become a random subsample of the full sample and there would be no non-response bias in the estimate (apart from a ratio estimate bias) when the data are further weighted by the inverse of the response rate. A slight ratio estimate bias may result because of the variation in the respondent sample. Since, in the majority of cases the probability of non-response is not known, every effort should be made to minimize the size of non-response. However, even if we did know the probability of non-response, there may still be response bias in the estimate based on the subsample just as would be present if there was no non-response.

Another major component of non-response is that of refusals and these can only be prevented, in many instances, by motivating them to respond. However, it is possible that those respondents who were initially reluctant to respond may commit larger response errors than those who were willing to co-operate so that while we have reduced the imputation error $^{NR}\varepsilon_i$, we may have increased the response error $^{R}\varepsilon_i$ (Platek, Singh and Tremblay [7]). Just to convert every refusal into a respondent may therefore lead to a false sense of security with respect to the validity of the responses. A well-trained interviewer will certainly succeed in motivating more refusals to respond and in obtaining more reliable responses than a poorly-trained interviewer.

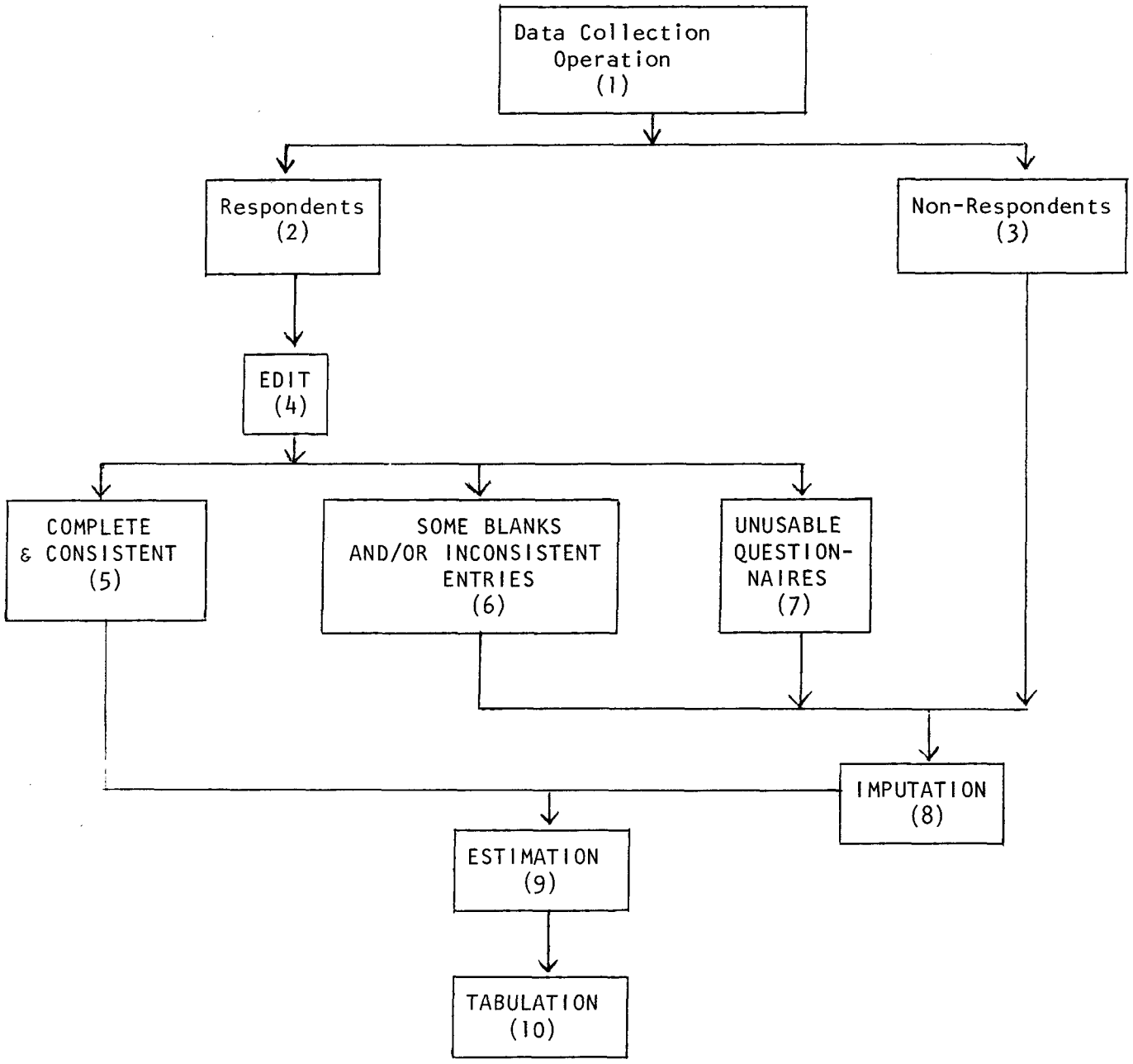
One method of dealing with non-response at the data collection stage is to substitute other previously unselected units in the field; for example, a next-door neighbour. Unfortunately, this would lead to a sampling bias. While any unit may be selected with known probability according to the sample design, substitution of other previously unselected respondents to replace unco-operative respondents in some uncontrolled manner, or even in a controlled manner, will alter the inclusion probabilities to such an extent that they cannot even be calculated. While a sampling bias of unknown magnitude would exist (since the selection probabilities are unknown for several reasons), the sampling variance may be reduced because of an increase in the effective sample size but there would probably be no reduction in the response error or the non-response bias. Even if the inclusion probabilities could be calculated, the non-response bias would remain since the unco-operative units essentially have no chance of inclusion.

In addition to callbacks or substitution of units in the field, interviewers may apply (i) double sampling (selecting a subsample of non-respondents and making an intensive effort to obtain responses from these units, or (ii) Politz scheme (considering "best time to call" as one of the weighting groups). These schemes are also expensive and must be carefully planned if they are to be used to tackle non-response.

3. TYPES OF RESPONSES AND CONCEPTUAL ISSUES OF IMPUTATION

As the information flows from data collection to tabulation, the various types of responses can be identified and are presented as follows in Chart 1.

Chart 1: Flow Chart Pertaining to Each Sampled Unit



This is, of course, a highly simplified diagram of the process and it is produced only for the purpose of the discussion of this paper.

Looking at Chart 1, two of the three groups following the edit stage require some action prior to estimation. These are the unusable questionnaires and the questionnaires containing some blanks and/or inconsistent entries. The unusable questionnaires can be classified as total non-response or they can be associated with the respondent households with some blank or inconsistent entries. There remain two groups that require some attention. The first group consists of blank and/or inconsistent responses, the second group consists of non-respondents. Non-respondents (at least in household surveys as opposed to the census) are usually weighted up in some manner. The deficient questionnaires, on the other hand, fall into two categories such as inconsistent entries or illegitimate blanks.

The inconsistent entries can be either logical impossibilities or they can be plausible but highly unlikely. It seems natural that if the entries are logical impossibilities and they can be detected as such, they ought to be adjusted even though they may not affect the data to any great extent. The adjustment would save a great deal of embarrassment on the part of subject matter analysts associated with the published reports.

In the case of plausible but highly unlikely entries, one is faced with a difficult choice between remaining with observations in an unnatural distribution or removing the extreme values of the distribution which may actually represent the real life situation. Ideally, one ought to opt for one or the other choice on the basis of experience with error mechanisms and the nature of the substantive distribution based on the knowledge of subject matter. In any case, one has to be able to identify the problem cases, i.e., one has to have suitable edit rules whenever one encounters impossible or highly unlikely events and a method of dealing with them.

There is a fundamental distinction between editing and imputation. Let us consider the set of all possible code combinations on a

questionnaire. Editing can be defined as the division of this set into two mutually exclusive subsets: those combinations which are judged acceptable and those which are unacceptable, the latter including questionnaires with invalid blanks and inconsistent entries. Thus, editing is basically a diagnosis and operationally it must be defined by a set of rules. Imputation, on the other hand, is more in the nature of a treatment of data, although the two clearly interact.

As far as editing is concerned, the detection of logically impossible entries and invalid blanks presents no conceptual problems and with respect to the detection of inconsistencies, there are a number of options available. For example, one can compare pairs of fields and decide that the two are inconsistent and hence, one of them has to be changed. One can continue this procedure by comparing some other pairs of fields (or three fields at a time). Having detected a particular inconsistency, one may either impute immediately one of the fields involved to make these fields consistent with one another, or else complete the entire edit process before imputation begins. However, by looking at two or three fields at a time, one does not take into account all the possibilities. For example, if one makes all combinations of, say two or three fields consistent with one another, it does not mean that the whole record will be consistent. A system which has been developed in Statistics Canada is based on the approach that identifies all inconsistencies before any corrective action is taken. Then, in the face of all known inconsistencies between the fields of the given record, together with all the logical impossibilities and invalid blanks, one decides which field or set of fields, if corrected, would remove all the inconsistencies in the whole record.

Having determined which fields are going to be changed, the next step is, of course, to carry out imputation for them. The simplest situation occurs when there is only one possible value which can be imputed for that field in such a way that after the imputation the record will be consistent. Sometimes, there may be more than one value which would

make the record consistent. If this is the case, one would choose a particular value which is more predominant in the field or more plausible. A good example of this kind can be found in the Labour Force Survey where in the fall to spring months, for 15 and 16 year old persons, if there is no Labour Force characteristic entered, one imputes that they are "attending school", although it is not at all impossible they do not attend school. So long as the proportion of such cases is sufficiently small, the effect of this will be a slight increase in bias. At the same time, there will be some reduction in variance and the added advantage of the reduction of complexity in imputing.

In other situations where one could reasonably impute a whole range of values, one needs some other criteria. One possible criterion would be to minimize the mean square error of the resulting estimates. The question, however, arises, the mean square error of which estimates? With the continuously increasing demand for micro data tabulated in a number of different and unforeseen ways one really does not know which mean square error one ought to minimize. Furthermore, one would not know all the kinds of aggregates to which a particular record may contribute in different kinds of tabulations. Consequently, one would like to use some other criterion which would produce the most appropriate entry for a field in a particular record in relation to the other information in the record. In other words, how can one best predict the value of one field on the basis of knowing the other fields on the record. A good example of this kind of imputation is the use of previous month's data in the Labour Force Survey: for a particular person, one could hardly find a better imputed value, particularly in those cases where demographic characteristics change slowly. If one does not have information based on the past, the best imputed value may be the result of some sort of regression equation. For some household surveys, however, the application of regression is somewhat restricted due to the qualitative nature of variables. Consequently, one may adopt as a reasonable criterion that

the distribution after imputation should remain as close as possible to the distribution prior to imputation with respect to marginal distributions or preferably, if it is possible, with respect to joint distributions of all the variables to be imputed.

In most cases, to impute for non-response at the micro level as opposed to some aggregate level is mainly justified because of the lack of advanced knowledge as to the kind of aggregates that will be produced from the micro data file. However, in some situations where one knows one can limit the tabulation requirements in advance, imputing at the individual level may not always be necessary. This notably applies to surveys based on area samples where the primary sampling units are not likely to be split up in any subsequent tabulations. In this case, one can hardly do better in terms of the mean square error of any of the possible aggregates that one will produce, but to impute the average of that primary sampling unit[1].

4. PROCESSING AND ESTIMATION

One of the most common procedures for accounting for non-response at the processing and estimation stage is that of the design-dependent balancing area, in which the weights are further inflated by the inverse of the response rate. In a balancing area b, an estimate of the characteristic total is given by $\hat{X}_b = \sum_{i=1}^{n_b} x_i/\Pi_i$, where x_i is the response, Π_i = the inclusion probability, and n_b is the sample size in the balancing areas. If only m_b units respond, then the weight Π_i^{-1} is further inflated by the inverse of the response rate, m_b/n_b, i.e. by the factor n_b/m_b.

The balancing areas should, preferably, be determined at the planning stage rather than at the processing stage and they could be individual strata, groups of strata, a province, primary sampling unit, or a

---

[1] While the weight adjustment at the PSU level is justified for complete non-response, it would be inappropriate for either partial non-response or fields whose entries have been rejected on account of editing. If one carried out weighting at the individual field level, one could not properly cross-tabulate the data since records would have more than one weight.

cluster. The choice of balancing area is quite crucial since the non-response rates and the bias may differ from area to area.

An important methodological problem, for example, is to determine an optimum or in some way appropriate size of balancing area where "appropriate" refers to a proper size to ensure a sufficient response rate in order to prevent excessive weights and at the same time ensure the advantages due to the measures of homogeneity to help reduce the non-response bias. It can readily be shown that weight inflation of all the records in a balancing area to compensate for non-respondents is equivalent to the substitution of the mean values of all the weighted respondents to each non-respondent in the area. If a characteristic has a high measure of homogeneity (increasing with decreasing size of area), then weighting (or substitution of mean value) in small areas vs. large areas would tend to result in mean values that are more similar to the actual characteristic value of the non-respondent than would be the case in larger areas. Thus, the non-response bias would tend to be lower in the case of small balancing areas than in the case of large balancing units. What about the variance? As balancing areas become smaller, the weight inflation becomes more unstable as the variation in response rates becomes more unstable among many small balancing areas as opposed to a few large balancing areas and the instability of the weight would tend to increase the variance. Clearly, there is some trade-off on the size of the balancing area between small areas to take advantage of the measure of homogeneity and large areas to ensure stability in the weight adjustments. The possible extreme values of the sizes of balancing areas are the whole sample at the upper end and a size of '1' at the lower end. However, in the latter case, one is faced with the problem of what should be done if that unit fails to respond. Instead of substitution of the mean value, one would have to resort to regression analysis or superpopulation models (an entirely different approach to substitution) or else employ historic values.

The choice of the size of balancing area depends not only on the measure of homogeneity but also on the sample design, the sample size and the response rate. Surveys with low response rates would require larger balancing units than those with high response rates. One could utilize small balancing areas and adopt some procedure of collapsing them until the response rate reaches some respectable level (not too much below the overall response rate) so as to minimize the instability of the weight. The collapsing of balancing areas however adds a complex dimension to the variance estimation since one would have to consider the probabilities of collapsing 1, 2, 3, 4, etc., balancing areas and the choice of 1, 2, 3 or 4 balancing areas. While such a procedure is undertaken in LFS, the need to collapse is infrequent enough not to warrant special treatment for variance estimation purposes. Consequently, if any variance calculations or analysis other than mere averages or totals are contemplated, balancing areas that are expected to be stable without much collapsing should be incorporated into the sample design. That is, the response rates should be sufficiently large with high probability to avoid the necessity of collapsing balancing areas if variance estimation is contemplated. This would discourage one from using small areas to balance for non-response.

Instead of weighting by the inverse of the response rate in a balancing area, one could duplicate a sufficient number of units among the m_b respondents to bring the apparent sample size up to the original level of n_b units. However, it can be shown that an additional variance component occurs over that incurred when simple weighting is applied and in the case of srs, the sampling variance is considered alone, the increase would be up to about 12%, depending upon the response rate (see Hansen et al [3]). The main advantage of duplication vs. weighting lies in ensuring that integral rather than fractional weights are applied to each record. In certain types of published data, e.g., the number of persons with some characteristic, integral weights would tend to avoid rounding errors when sub-classifying data. When one estimates means or proportions or certain types of quantitative totals such as gross national product, the use of integers rather than fractional weights are of no advantage.

Apart from the comments in the above paragraph, the methodological problems concerned with weighting in balancing areas also apply to duplication in balancing areas.

Another important method of imputation for non-response is one of substitution of historical (previous month's data) or external source data (administrative files, other surveys, Census data). Once the substitution of historical or external source data has been undertaken to the extent possible for non-respondents, the weighting or duplication may be affected within balancing areas. In the case of weighting, one would inflate the weight Π_i^{-1} by the factor $n_b/(m_b+m_b')$, where m_b respondents were obtained as before and for m_b' of the (n_b-m_b) non-respondents, historical records were substituted for the missing data. In such a method of imputation, the sampling variance is reduced from that which occurs in the weighting scheme since we have increased the effective sample size from m_b to somewhere between m_b and (m_b+m_b') units. The increased sample, including those records imputed from historical records will not be as good as m_b+m_b' since historical or external source data are not as good as current response information unless there has been no change in the characteristics of the units for which substitution of historical data was undertaken.

Alternatively, one may wish to duplicate respondents; i.e., take a sample from respondents equal in size to the number of non-respondents and apply a weight of 2 instead of inflate the weight for all the respondents. However, one may wish to avoid duplication of those non-respondents for which substitution of historical information had been undertaken but one would rather subsample $n_b - (m_b+m_b')$, say m_b^* units from the m_b respondents to duplicate in order to bring the apparent sample size from (m_b+m_b') to n_b units in balancing area b. The estimated total for balancing area b would be

$$\hat{X}_b = \sum_{i=1}^{m_b} w_i\, x_i/\Pi_i + \sum_{j=m_b+1}^{m_b+m_b'} x_j'/\Pi_j, \qquad (4.1)$$

where x_j' is the imputed value for unit j and w_i = 1 or 2 (2 for a random subsample of $n_b - m_n - m_b'$ units among the m_b respondents). The expected value of $\hat{X}_b$ over all possible ways of duplicating is

$$\hat{X}_b^* = (n_b - m_b')/m_b \sum_{i=1}^{m_b} x_i/\Pi_i + \sum_{j=m_{b+1}}^{m_b+m_b'} x_j'/\Pi_j. \qquad (4.2)$$

Consequently, $V(\hat{X}_b) = V(\hat{X}_b^*)$ + (additional component of variance as a result of subsampling among the respondents). $\hat{X}_b^*$ is not the same as the estimate $n_b/(m_b+m_b')[\sum_{i=1}^{m_b} x_i/\Pi_i + \sum_{j=m_{b+1}}^{m_b+m_b'} x_j'/\Pi_j]$ and the variance of $\hat{X}_b^*$ is also different from that of the estimate where the weight inflation of $n_b/(m_b+m_b')$ is applied (see appendix).

The estimation procedures dealt with above include weighting or duplicating in design-dependent balancing areas. If historical or external source data are available for some of the non-respondents, these could be employed for imputation purposes prior to weighting or duplication in balancing areas. Instead of balancing areas, one could utilize weighting classes for imputation purposes and these are discussed in the next paragraph.

Weighting classes are distinguished from balancing areas in that they generally comprise characteristics of ultimate units (e.g., dwelling types, special income groups, etc.) as opposed to geographic areas, though one could conceivably group areas according to some distinct characteristics that are not related to the sample design. Usually, one defines weighting classes as well as balancing areas prior to the survey gathering procedure and makes adjustments through collapsing if the response rates are unacceptably low or the sample too small to employ any type of adjustment of the weights. In some imputation procedures, however, weighting classes are defined after the survey data have been gathered where factor analysis or other analytical tools are employed

to determine the most efficient set of weighting classes. After the weighting classes have been determined, the estimation procedures are essentially identical to those used in balancing areas. The biases and the variances (at least in terms of individual and joint inclusion probabilities of the ultimate units and other parameters not related to the sample design) are identical. However, upon further expansion of the variance to take into account the particular sample design, the variances of the estimates pertaining to balancing areas and weighting classes will be quite different. In order to utilize weighting classes for imputation purposes some knowledge about the non-respondents (such as income class, size of household dwelling type) must be available. In practice, when such information is not available, the procedure cannot be used.

The estimation formula for the methods of imputation discussed here may be written as below.

$$\hat{X} = \sum_{b} \hat{X}_b \quad \text{estimates the total of some characteristic,}$$

where b is either the balancing area or weighting class. The estimate for a given balancing area or weighting class is in turn given by:

$$\hat{X}_b = \sum_{i=1}^{m_b} w_i x_i/\Pi_i + \sum_{j=m_b+1}^{m_b+m_b'} w_j x_j'/\Pi_j, \text{ where } \Pi_i \text{ or } \Pi_j \text{ is the} \quad (4.3)$$

inclusion probability and m_b is the number of units that responded out of n_b units in balancing area b.

x_i = response value for unit $i = X_i$ (true value) + $^{R}\varepsilon_i$ (response error) (4.4)

x_j' = historical value for unit j (if available), given that unit j failed to respond. Among the $(n_b - m_b)$ non-responding units m_b' possess historical records in balancing area b.

x_j' = X_j (true value) + $^{R}\varepsilon_j'$ (response error of historical value, relative to X_j) (4.5)

w_i and w_j are weights, appropriate to the imputation method and the weights are listed in Table 1.

Table 1: Imputation Method in Balancing Area/Weighting Class

		w_i	w_j
(a)/(c)	Weighting by Inverse of Response Rate m_b/n_b	n_b/m_b	0
(b)/(d)	Duplication of a random subsample of (n_b-m_b) units from m_b respondents	2 for (n_b-m_b) units & 1 for $(2m_b-n_b)$ units	0
(e1)	Substitution of Historical Records for m_b' of (n_b-m_b) non-respondents, followed by weighting	$n_b/(m_n+m_b')$	$n_b/(m_b+m_b')$
(e2)	Substitution as in (iii), followed by duplication of respondents only	2 for $(n_b-m_b-m_b')$ units & 1 for $(2m_b+m_b'-n_b)$ units	1

(c) and (d) refer to weighting classes while (a) and (b) refer to balancing areas.

In the case of duplication, we have assumed the response rate m_b/n_b to be at least 0.5. If it is exactly 0.5, then duplication and weighting would yield identical estimates. Let us suppose that $n_b/m_b = W_b+d_b$, where W_b is an integer and d_b, a fraction in the range $0 \leq d_b < 1$, then m_b would be partitioned into m_{b1} units, subsampled at random requiring a weight of W_b and $m_{b2} = (m_b-m_{b1})$ units, requiring a weight of (W_b+1). Thus, $n_b = W_b m_b+d_b m_b = W_b m_{b1}+(W_b+1)\ m_{b2} = W_b m_b+m_{b2}$. Hence, $m_{b2} = d_b m_b$ and $m_{b1} = (1-d_b)\ m_b$. Consequently, a random subsample of $d_b m_b$ respondents would be assigned a weight of (W_b+1) and the remaining $(1-d_b)\ m_b$ respondents assigned a weight of W_b. If $W_b = 1$, then $n_b = m_b+d_b m_b$ or $d_b m_b = (n_b-m_b)$ units would require a weight of 2, as indicated in Table 1. Whatever the value of W_b, the expected value of the estimates

by method (b) or (d) over all possible subsamples of $d_b m_b$ respondents which would be assigned a weight of (W_b+1) instead of W_b is just the estimate by the weight inflation as of method (a) or (c).

In the case of method (e2), use of historical or external source data, followed by duplication of respondents, one would most likely confine the duplication only to a subsample from the m_b respondents rather than from the (m_b+m_b') units that either responded or utilized historical records. In such a case, the conditional expected value over all possible random subsamples of units assigned for duplication, given the sample, is not the estimate by method (e1) but rather an estimate with $w_i = (n_b-m_b')/m_b$ for the m_b respondents and $w_j = 1$ for the m_b' non-respondents with available historical records.

5. BIAS OF ESTIMATES

The bias of $\hat{X}_b$ according to imputation procedure may be readily obtained simply by finding $E\,\hat{X}_b$. Since $\hat{X}_b$ is a ratio estimate with the responding sample m_b a variable and similarly for (m_b+m_b'), a ratio estimate bias exists in addition to the response and non-response biases but we have neglected this in Table 2 where the biases are given for the estimates which are defined in Table 1. In the table, α_i is the probability of unit i responding while $\bar{\alpha}_b$ is the expected response rate in balancing area b and may be written as $\bar{\alpha}_b = E_b^* \alpha_i$. ${}^{R}B_i$ denotes the response bias pertinent to unit i while ${}^{R}B_i'$ denotes the bias of the historical value, relative to the true value X_i. α_i' is the probability of unit i possessing historical data and finally, Cov $\delta_i\delta_i'$ is the covariance between the event of responding or not responding (δ_i = 1 or 0) and the existence or non-existence of historical data (δ_i' = 1 or 0).

It will be noted in Table 2 that the bias is identical for weighting and duplicating and the reason for this is that, as pointed out before, the expected value of the estimate using duplication for imputation purposes over all possible subsamples of units to be duplicated is just the estimate using the weight inflation. The overall expected value of the two estimates is consequently the same.

The bias under method (e1) may be readily compared with the bias under methods (a) to (d). The non-response bias under method (e1) is given by $(\bar{\alpha}_b + \bar{\alpha}_b'')^{-1}$ En_b Cov_b^* $(\alpha_i + \alpha_i'', X_i/\Pi_i)$, which will reduce to the non-response bias according to methods (a) to (d) when $\alpha_i'' = 0$ and $\bar{\alpha}_b'' = 0$. As the combined probabilities $\alpha_i + \alpha_i''$ approach one, the population covariance between $\alpha_i + \alpha_i''$ and X_i/Π_i or Cov_b^* $(\alpha_i + \alpha_i'', X_i/\Pi_i)$ approaches zero. In fact, if $\alpha_i + \alpha_i''$ were equal for all i, the covariance would be zero and there would be no non-response bias. The same holds for methods (a) to (d) if α_i's were all equal. The non-response bias under method (e1) would be expected to be lower than under methods (a) to (d) because of an anticipated decrease in the population covariance. Depending upon the extent of the availability of historical records, $\alpha_i + \alpha_i''$ would exceed α_i and would most likely have a smaller population variance. If Cov_b^* $(\alpha_i, X_i/\Pi_i) = r_{\alpha_i, X_i/\Pi_i} \sqrt{V_b^*(\alpha_i)\, V_b^*(X_i/\Pi_i)}$ and if Cov_b^* $(\alpha_i+\alpha_i', X_i/\Pi_i) = r_{\alpha_i+\alpha_i', X_i/\Pi_i} \sqrt{V_b^*(\alpha_i+\alpha_i')} \cdot \sqrt{V_b^*(X_i/\Pi_i)}$, then Cov_b^* $(\alpha_i+\alpha_i', X_i/\Pi_i)$ would most likely be smaller than Cov_b^* $(\alpha_i, X_i/\Pi_i)$ because one would expect $(\alpha_i + \alpha_i')$'s to be closer to one and presumably less variable among the units than α_i's alone, implying that $V_b^*(\alpha_i + \alpha_i') < V_b^*(\alpha_i)$. A further decrease in the non-response bias would occur under methods (e1) than under methods (a) to (d) because of the larger denominator $(\bar{\alpha}_b + \bar{\alpha}_b'')$ pertaining to method (e1) compared with $\bar{\alpha}_b$ in the denominator of the bias pertaining to methods (a) to (d).

A lower non-response bias may be partially offset by a larger response bias pertaining to method (e1). If ${}^{R}B_i'$'s were about the same magnitude as ${}^{R}B_i$'s on an average, then the response biases would be about the same but one would expect ${}^{R}B_i'$'s to be slightly larger than ${}^{R}B_i$'s since historical data would not be as close to the truth as current responses.

Table 2: Bias of Estimate, According to Imputation Procedure[1]

Method	Bias of Estimate	
Weighting (a)/(c) & Duplicating (b)/(d)	$\bar{\alpha}_b^{-1} En_b \, Cov_b^* (\alpha_i, X_i/\Pi_i)$	... non-response bias
	$+ \bar{\alpha}_b^{-1} \sum_{i=1}^{N_b} \alpha_i \, {}^{R}B_i$	... response bias
(e1) Substitution of Historical Records, then weighting	$(\bar{\alpha}_b + \bar{\alpha}_b'')^{-1} En_b \, Cov_b^*(\alpha_i, X_i/\Pi_i)$	... non-response bias contributed by the use of weight inflation of respondents
	$+ (\bar{\alpha}_b + \bar{\alpha}_b'')^{-1} En_b \, Cov_b^*(\alpha_i'', X_i/\Pi_i)$	... non-response bias, contributed by substitution of historical records
	$+ (\bar{\alpha}_b + \bar{\alpha}_b'')^{-1} \sum_{i=1}^{N_b} (\alpha_i \, {}^{R}B_i + \alpha_i'' \, {}^{R}B_i')$	... response bias, contributed respectively by respondents and by non-respondents with historical data
(e2) Substitution of Historical Records, then duplication or weighting of respondents only	$\bar{\alpha}_b^{-1} En_b (1-\bar{\alpha}_b'') Cov_b^*(\alpha_i, X_i/\Pi_i)$	... non-response bias contributed by duplication
	$+ \bar{\alpha}_b^{-1} En_b (1-\bar{\alpha}_b'') \sum_{i=1}^{N_b} \alpha_i \, {}^{R}B_i$	... response bias contributed by respondents
	$+ En_b \, Cov_b^*(\alpha_i'', X_i/\Pi_i)$	... non-response bias, contributed by substitution of historical records
	$+ \sum_{i=1}^{N_b} \alpha_i'' \, {}^{R}B_i'$	... response bias, contributed by imputation by historical records

In (e1) and (e2), $\alpha_i'' = (1-\alpha_i)\alpha_i' - Cov \, \delta_i \, \delta_i'$ and $\bar{\alpha}_b'' = E_b^* \alpha_i'' = (1-\bar{\alpha}_b) \, \bar{\alpha}_b' - Cov_b^* \alpha_i \alpha_i' - E_b^* Cov \delta_i \delta_i'$

[1] Bias derived for method (e1) in Appendix 1.

6. VARIANCE OF ESTIMATES

The variance of $\hat{X}_b$ as defined for method (e1) is partially derived in Appendix 2 by regarding $\hat{X}_b$ as a combined product and ratio expression and employing Tayler series expansions. The same holds for $Cov(\hat{X}_b, \hat{X}_c)$,

$$V(\hat{X}_b) \doteq \{X_b + (\bar{\alpha}_b + \bar{\alpha}_b'')^{-1}[En_b \, Cov_b^*(\alpha_i + \alpha_i'', X_i/\Pi_i) + \sum_{i=1}^{N_b}(\alpha_i \, {}^R B_i + \alpha_i'' \, {}^R B_i')]\}^2$$

$$\times \{V_s[\frac{n_b}{En_b} + \frac{\sum_{i=1}^{n_b} \Pi_i^{-1}[\alpha_i(X_i + {}^R B_i) + \alpha_i''(X_i + {}^R B_i')]}{\sum_{i=1}^{N_b}[\alpha_i(X_i + {}^R B_i) + \alpha_i''(X_i + {}^R B_i')]}$$

$$- \frac{\sum_{i=1}^{n_b}(\alpha_i + \alpha_i'')}{En_b(\bar{\alpha}_b + \bar{\alpha}_b'')}] \quad \ldots \text{ sampling variance (s referring to a specific sample)}$$

$$+ E_s \, V[\frac{\sum_{i=1}^{n_b} \Pi_i^{-1}[\delta_i(X_i + {}^R \varepsilon_i) + \delta_i''(X_i + {}^R \varepsilon_i')}{\sum_{i=1}^{N_b}[\alpha_i(X_i + {}^R B_i) + \alpha_i''(X_i + {}^R B_i')]}$$

$$- \frac{\sum_{i=1}^{n_b}(\delta_i + \delta_i'')}{En_b(\bar{\alpha}_b + \bar{\alpha}_b'')}] | s \quad \ldots\ldots \text{ non-sampling variance } \} \qquad (6.1)$$

where expanded forms of non-sampling variances and covariances may be obtained in Appendix 2. Similarly $Cov(\hat{X}_b, \hat{X}_c)$ may be expressed. In the above formula, $\delta_i'' = (1-\delta_i)\delta_i'$ and $\alpha_i'' = E\delta_i''$.

In the case of methods (a) and (c), formula 6.1 also holds with all α_i'', δ_i'' and $\bar{\alpha}_b''$ equal to zero.

For methods (b) and (d); viz., duplication of a subsample of units at random to boost the sample from m_b to n_b units in balancing area b, formula 6.1 yields one component of variance. There is an additional component, arising from the variation in the choice of subsampled units to be duplicated.

The additional variance component is given by:

$$E_s \, E\Big[\sum_{i=1}^{n_b} (w_i - \frac{n_b}{m_b}) \, \delta_i (X_i + {}^{R}\varepsilon_i) \, \Pi_i^{-1} \Big]^2 \Big| s, \qquad (6.2)$$

where s is a given sample of n_b units and the second E is taken over all possible responses and non-responses within a particular sample. For a given response rate m_b/n_b, $E\, w_i = n_b/m_b$ for all respondents in balancing area b and the response rate is assumed to be at least 0.5 so that $w_i = 1$ or 2. In the case of srswor, assuming m_b and n_b both constant, Hansen et al [3] showed that the additional variance component caused by duplication instead of weighting is as much as 12% for a response rate of about 3/4. Similar results occur when ppswor is undertaken. However, when m_b and n_b both vary, further studies on the expansion of 6.2 must be carried out.

It is difficult to compare the variance of $\hat{X}_b$ under method (e2) with that under method (a) or (c) from formula 6.1 without substitution of numerical values. Intuitively, one would expect the variance under method (e1) to be lower than that under method (a), the extent of the decrease depending upon the size of the non-response utilizing historical records and the correlation between historical and current information. The variances need to be explored, perhaps upon rewriting 6.1 in terms of average parametric values of α_i, ${}^{R}\sigma_i^2$, α_i'', etc. in the balancing area.

7. CONCLUSION

The conceptual issues cover the difficulty of non-response and pros and cons of different methods of dealing with them. Empirical data will be needed to obtain the parameters in the formulae stated in this paper for comparison purposes. An important fact to be noted is the additional variance component that occurs in duplication as opposed to weighting when a given response rate occurs in a given sample. The effect of duplication must be further studied as sample size and response rates both vary.

Much of the methodological development of the bias and the variance of estimates under different imputation procedures depends upon the knowledge of response probabilities which are rarely known in real life. Some estimates of response probabilities can be obtained from longitudinal studies of response profiles in the case of continuous surveys; otherwise, special experimental studies of non-respondents outside the sample used for publication purposes may be needed to obtain approximate estimates of response probabilities.

It is very important to note that, under the usual imputation procedures of duplication or weighting, there is non-response bias only if the response probabilities vary among the units and if there exists a correlation between response probabilities and the characteristic values of the units. Response bias, however, will occur whether or not we have full response.

8. ACKNOWLEDGMENT

The authors sincerely appreciate the comments and suggestions of the referee, editor, and A. Ashraf, Senior Methodologist, Household Surveys Development Division.

RESUME

L'article analyse les problèmes posés par les mesures applicables, à diverses étapes de la planification d'une enquête, pour contrer la non-réponse, les répercussions de ces mesures sur l'écart-type moyen, ainsi que l'utilité pratique, les avantages et les inconvénients de ces mesures. Il examine aussi certaines questions théoriques touchant la complexité et les niveaux d'imputation. Il existe diverses méthodes d'imputation: par pondération, par reproduction et par substitution d'enregistrements. L'article traite aussi de certaines questions méthodologiques concernant le biais et la variance.

REFERENCES

[1] Fellegi, I.P. and Holt, D., "A Systematic Approach to Automatic Edit and Imputation", Journal of the American Statistical Association (1976), Vol. 71, pp. 17-35.

[2] Ghangurde, P.D. and Mulvihill, J., "Non-Response and Imputation in Longitudinal Estimation in LFS", Household Surveys Development Staff, Statistics Canada Report (February 1978).

[3] Hansen, M.H., Hurwitz, W.N. and Madow, W.G., "Sample Survey Methods and Theory", Vol. II, Theory, pp. 139-141, John Wiley and Sons, Inc. (1953).

[4] Nordbotten, S., "The Efficiency of Automatic Detection and Correction of Errors in Individual Observations as Compared with Other Means for Improving the Quality of Statistics", Bulletin of the International Statistical Institute, Proceedings of the 35th Session, Belgrade 41, (September 1965), pp. 417-441.

[5] Platek R., "Imputation for Household Surveys in Statistics Canada", report prepared for European Statisticians' Conference held in Geneva, March 1978.

[6] Platek, R., "Some Factors affecting Non-Response", Survey Methodology (Statistical Services Field, Statistics Canada), Vol. 3, No. 2 (December 1977), pp. 191-214.

[7] Platek, R., Singh, M.P. and Tremblay, V., "Adjustment for Non-Response in Surveys", Survey Methodology (Statistical Services Field, Statistics Canada), Vol. 3, No. 1 (June 1977), pp. 1-24.

[8] Platek, R. and Gray, G.B., "Imputation Methodology", Household Surveys Development Division, technical paper describing biases and variances of estimates, using different methods of imputation.

[9] Szameitat, K. and Zindler, H.J., "The Reduction of Errors in Statistics by Automatic Corrections", Bulletin of the International Statistical Institute, Proceedings of 35th Session, Belgrade 41, (September 1965), pp. 395-417.

APPENDIX 1

Bias of Estimate in Balancing Area/Weighting Class

Consider the estimate $\hat{X}_b$ as defined by 4.3 in general, and in particular for case (e1) as of Table 1, viz., substitution of historical records for m_b' of $(n_b - m_b)$ non-respondents, followed by weighting.

$$\text{Then} \qquad \hat{X}_b = \frac{n_b}{m_b + m_b'} \left[\sum_{i=1}^{m_b} x_i/\Pi_i + \sum_{j=m_b+1}^{m_b+m_b'} x_j'/\Pi_j \right] \qquad \text{A1.1}$$

To derive the bias of $\hat{X}_b$, let us define δ_i as 1 or 0 according as unit i responds or not and δ_i' = 1 or 0 according as historical records are available and used for imputation or not. Then $m_b = \sum_{i=1}^{n_b} \delta_i$ and $m_b' = \sum_{i=1}^{n_b} (1-\delta_i)\,\delta_i'$. In the case of methods (a) to (d) all $\delta_i' = 0$ and consequently $m_b' = 0$.

$$\text{Then} \qquad \hat{X}_b = \frac{n_b}{\sum_{i=1}^{n_b} [(\delta_i + (1-\delta_i)\delta_i']} \sum_{i=1}^{n_b} \Pi_i^{-1} [\delta_i (X_i + {}^{R}\varepsilon_i) + (1-\delta_i)\delta_i' (X_i + {}^{R}\varepsilon_i')] \qquad \text{A1.2}$$

in which x_i and x_j' defined by 4.4 and 4.5 respectively have been substituted.

To determine the bias, one needs only to derive $E\hat{X}_b$ as of A1.2. We shall neglect the ratio estimate bias and also the covariance between n_b and the ratio with $\sum_{i=1}^{n_b}$ in the numerator and denominator, a covariance which may exist when the sample size, n_b, is a variable.

$$\text{Then} \quad \hat{EX}_b \doteq \frac{En_b \sum_{i=1}^{N_b} [\alpha_i (X_i + {}^R B_i) + (\alpha_i' - \alpha_i \alpha_i' - \text{Cov } \delta_i \delta_i')(X_i + {}^R B_i')]}{\sum_{i=1}^{N_b} \Pi_i [\alpha_i + (1-\alpha_i)\alpha_i' - \text{Cov } \delta_i \delta_i']} \qquad \text{A1.3}$$

noting that $E^R \varepsilon_i = {}^R B_i$ and $E^R \varepsilon_i' = {}^R B_i'$.

We have not assumed independence between δ_i and δ_i' since the presence of historical record may be related to the tendency to respond or not to respond. Hence, $E(1-\delta_i)\ \delta_i' = (1-\alpha_i)\alpha_i' - \text{Cov } \delta_i \delta_i'$.

Further simplification of A1.3 is possible by utilizing "average" parameters such as, for example, $E_b^* T_i = \sum_{i=1}^{N_b} (\Pi_i / En_b)\ T_i = \bar{T}_b$, whatever T_i may be. Other expressions such as $\text{Cov}_b^* (T_i U_i) = E_b^* T_i U_i - E_b^* T_i E_b^* U_i$ and $V_b^*(T_i) = \text{Cov}_b^*(T_i, T_i)$ may be derived. E_b^* is a weighted average of individual parameter values, using Π_i / En_b as the weights, noting that $\sum_{i=1}^{N_b} \Pi_i = En_b$.

$$\text{Thus,} \quad \sum_{i=1}^{N_b} \alpha_i X_i = \sum_{i=1}^{N_b} \Pi_i \alpha_i\ X_i / \Pi_i = En_b\ E_b^* \alpha_i\ X_i / \Pi_i$$

$$= En_b [E_b^*\ \alpha_i\ E_b^*\ X_i / \Pi_i + \text{Cov}_b^* (\alpha_i,\ X_i / \Pi_i)]$$

$$= En_b [\bar{\alpha}_b (En_b)^{-1}\ X_b + \text{Cov}_b^* (\alpha_i,\ X_i / \Pi_i)]$$

$$= \bar{\alpha}_b\ X_b + En_b\ \text{Cov}_b^* (\alpha_i,\ X_i / \Pi_i) \qquad \text{A1.4}$$

Now let $(1-\alpha_i)\alpha_i' - \mathrm{Cov}\delta_i\delta_i' = \alpha_i''$ A1.4a

$$E_b^* \alpha_i'' = (1-\bar{\alpha}_b)\bar{\alpha}_b' - \mathrm{Cov}_b^*\alpha_i,\alpha_i' - E_b^* \mathrm{Cov}\delta_i\delta_i' = \alpha_b'' \text{ say} \qquad \text{A1.5}$$

In a similar manner as undertaken in A1.4,

$$\sum_{i=1}^{N_b} \alpha_i'' X_i = \bar{\alpha}_b'' X_b + En_b \mathrm{Cov}_b^*(\alpha_i'', X_i/\Pi_i), \text{ where } E_b^*\alpha_i'' = \bar{\alpha}_b''.$$

Then $\hat{EX}_b \doteq En_b[(\bar{\alpha}_b+\bar{\alpha}_b'') X_b + En_b \mathrm{Cov}_b^*(\alpha_i+\alpha_i'', X_i/\Pi_i)$

$$+ \sum_{i=1}^{N_b} (\alpha_i R_{B_i} + \alpha_i'' R_{B_i}')]/En_b(\bar{\alpha}_b+\bar{\alpha}_b''), \text{ where the bias equals}$$

$$(\bar{\alpha}_b+\bar{\alpha}_b'')^{-1}[En_b \mathrm{Cov}_b^*(\alpha_i+\alpha_i'', X_i/\Pi_i) + \sum_{i=1}^{N_b} (\alpha_i R_{B_i} + \alpha_i'' R_{B_i}')] \qquad \text{A1.6}$$

The bias under imputation methods (a) to (d) immediately follow by putting $\alpha_i'' = 0$ and $\bar{\alpha}_b'' = 0$ in A1.6.

The bias of $\hat{X}_b$ under method (e2) may be similarly derived except that the covariance between δ_i and δ_j'' has been omitted in order to simplify the formula.

APPENDIX 2

Summary of Development of Variance of Estimate $\hat{X}$

$$\hat{X} = \sum_b \hat{X}_b, \text{ where } b = \text{balancing area or weighting class}$$

$$V(\hat{X}) = \sum_b V(\hat{X}_b) + \sum_{b \neq c} Cov(\hat{X}_b, \hat{X}_c), \text{ and hence, the need to derive}$$

$V(\hat{X}_b)$ and $Cov(\hat{X}_b, \hat{X}_c)$. A covariance may exist between the estimates based on different balancing areas or weighting classes, depending upon the definition of balancing areas or weighting classes as well as the sample design.

We will deal with method (e1), where historical information is substituted for non-responses, whenever available or appropriate and then weighting or duplication of records to boost the sample up to the required level. We will deal with the weighting first, where $\hat{X}_b$ is defined as in 4.3 or A1.1 or A1.2 (the most convenient form for the development of the bias and variance).

$\hat{X}_b$ may be regarded as a complex expression of the form $n_b y_b / z_b$ with n_b, y_b and z_b all variables. In some sample designs, n_b remains constant but it need not be in the general developments.

Then, by the use of Taylor series expansion,

$$Cov(n_b y_b/z_b,\ n_c y_c/z_c) \doteq (En_b\ Ey_b/Ez_b)(En_c\ Ey_c/Ez_c)$$

$$[\text{Rel Cov } n_b,n_c + \text{Rel Cov } y_b,y_c + \text{Rel Cov } z_b,z_c$$

$$+ (\text{Rel Cov } n_b,y_c + \text{Rel Cov } n_c,y_b) - (\text{Rel Cov } n_b,z_c + \text{Rel Cov } n_c,z_b)$$

$$- (\text{Rel Cov } y_b,\ z_c + \text{Rel Cov } y_c,\ z_b)] \qquad \text{A2.1}$$

and Rel Var $(n_b y_b/z_b)$ immediately follows by putting c=b.

To derive $V(\hat{X}_b)$ and $Cov(\hat{X}_b,\hat{X}_c)$, we require the following expressions. Here $(1-\delta_i)\delta_i$ as defined in A1.2 will be abbreviated by δ_i'' so that $E\delta_i''/i = \alpha_i''$ as defined in A1.4a.

En_b cannot be further simplified than $\sum_{i=1}^{N_b} \Pi_i$

$$V(n_b) = \sum_{i\neq j}^{N_b} \Pi_{ij} - En_b(En_b-1)$$

Ey_b is given by [ ] in A1.6 and $Ez_b = En_b(\bar{\alpha}_b+\bar{\alpha}_b'')$, with $\bar{\alpha}_b''$ given in A1.5

Additional expressions involve variances and covariances, which are stated below without proof but have been developed by Platek and Gray [8].

$$V \sum_{i=1}^{n_b} \Pi_i^{-1}[\delta_i(X_i+{}^R\varepsilon_i) + \delta_i''(X_i+{}^R\varepsilon_i')]$$

$$= V_s\{ \sum_{i=1}^{n_b} \Pi_i^{-1}[\alpha_i(X_i+{}^RB_i) + \alpha_i''(X_i+{}^RB_i')]\}$$

$$+ E_s\{ V \sum_{i=1}^{n_b} \Pi_i^{-1}[\delta_i(X_i+{}^R\varepsilon_i) + \delta_i''(X_i+{}^R\varepsilon_i')]|s \ , \qquad \text{A2.2}$$

where s means a specific sample of n_b units.

The second line, viz., the non-sampling variance component is given by:

$$\sum_{i=1}^{N_b} V[\delta_i(X_i+{}^R\varepsilon_i) + \delta_i''(X_i+{}^R\varepsilon_i')]\Pi_i^{-1}$$

$$+ \sum_{i\neq j}^{N_b} \Pi_{ij}\Pi_i^{-1}\Pi_j^{-1} \ Cov[\delta_i(X_i+{}^R\varepsilon_i) + \delta_i''(X_i+{}^R\varepsilon_i'),\delta_j(X_j+{}^R\varepsilon_j)+\delta_i''(X_j+{}^R\varepsilon_j')].$$

Here, $V[\delta_i(X_i+{}^R\varepsilon_i)+\delta_i''(X_i+{}^R\varepsilon_i')] = \alpha_i\ {}^R\sigma_i^2 + \alpha_i''\ {}^R\sigma_i'^2$

$$+ 2(\alpha_i\alpha_i''+ \mathrm{Cov}\delta_i\delta_i'')r_{2ii''}\ {}^R\sigma_i\ {}^R\sigma_i'+V[\delta_i(X_i+{}^RB_i)+\delta_i''(X_i+{}^RB_i')]$$

and $\mathrm{Cov}[\delta_i(X_i+{}^R\varepsilon_i)+\delta_i''(X_i+{}^R\varepsilon_i'),\delta_j(X_j+{}^R\varepsilon_j)+\delta_j''(X_j+{}^R\varepsilon_j')]$

$$= (\alpha_i\alpha_j+\mathrm{Cov}\delta_i\delta_j)\ r_{2ij}\ {}^R\sigma_i\ {}^R\sigma_j+(\alpha_i\alpha_j''+\mathrm{Cov}\delta_i\delta_j'')\ r_{2ij}''\ {}^R\sigma_i\ {}^R\sigma_j'$$

$$+ (\alpha_i''\alpha_j+\mathrm{Cov}\delta_i''\delta_j)\ r_{2ji}''\ {}^R\sigma_i'\ {}^R\sigma_j+(\alpha_i''\alpha_j''+\mathrm{Cov}\delta_i''\delta_j'')\ r_{2ij}'\ {}^R\sigma_i'\ {}^R\sigma_j'$$

$$+ \mathrm{Cov}[\delta_i(X_i+{}^RB_i)+\delta_i''(X_i+{}^RB_i'),\delta_j(X_j+{}^RB_j)+\delta_j''(X_j+{}^RB_j')].$$

In the above, $\mathrm{Cov}\ {}^R\varepsilon_i\ {}^R\varepsilon_j = r_{2ij}\ {}^R\sigma_i\ {}^R\sigma_j$ = covariance between current responses of pairs of units, $\mathrm{Cov}\ {}^R\varepsilon_i\ {}^R\varepsilon_j' = r_{2ij}''\ {}^R\sigma_i\ {}^R\sigma_j'$ = covariance between current and historical responses (applicable also for j=i), and $\mathrm{Cov}\ {}^R\varepsilon_i'\ {}^R\varepsilon_j' = r_{2ij}'\ {}^R\sigma_i'\ {}^R\sigma_j'$ = covariance between historical responses.

If we replace α_i by $\alpha_i^2+V(\delta_i)$ and similarly for α_i'', the above variances and covariances would be symmetrically described.

$$V\left[\sum_{i=1}^{n_b}(\delta_i+\delta_i'')\right]$$

$$= V_s\left[\sum_{i=1}^{n_b}(\alpha_i+\alpha_i'')\right]$$

$$+ E_s\ V\sum_{i=1}^{n_b}(\alpha_i+\alpha_i'')\,|\,s$$

$$= V_s\sum_{i=1}^{n_b}(\alpha_i+\alpha_i'') + \sum_{i=1}^{N_b}\Pi_i[V(\delta_i+\delta_i'')]$$

$$+ \sum_{i\neq j}^{N_b}\Pi_{ij}[\mathrm{Cov}(\delta_i+\delta_i''),(\delta_j+\delta_j'')]\ . \qquad \text{A2.4}$$

$$\text{Cov} \quad \sum_{i=1}^{n_b} (\delta_i+\delta_i'') , \quad \sum_{i=1}^{n_b} \Pi_i^{-1} \left[\delta_i(X_i+{}^R\varepsilon_i)+\delta_i''(X_i+{}^R\varepsilon_i')\right]$$

$$= \text{Cov} \left\{ \sum_{i=1}^{n_b} (\alpha_i+\alpha_i'') , \quad \sum_{i=1}^{n_b} \Pi_i^{-1} \left[\alpha_i(X_i+{}^RB_i)+\alpha_i''(X_i+{}^RB_i')\right]\right\}$$

$$+ E_s \text{ Cov}\left\{ \sum_{i=1}^{n_b} (\delta_i+\delta_i'') , \quad \sum_{i=1}^{n_b} \Pi_i^{-1} \left[\delta_i(X_i+{}^R\varepsilon_i)+\delta_i''(X_i+{}^R\varepsilon_i')\right]\right\} | s \qquad \text{A2.5}$$

The second line, viz., the non-sampling covariance is given by:

$$\sum_{i=1}^{N_b} \text{Cov}(\delta_i+\delta_i'') , \quad \left[\delta_i(X_i+{}^RB_i)+\delta_i''(X_i+{}^RB_i')\right] / i$$

$$+ \sum_{i\neq j}^{N_b} \Pi_{ij} \Pi^{-1} , \quad \text{Cov}(\delta_i+\delta_i'') \cdot \left[\delta_j(X_j+{}^RB_j)+\delta_j''(X_j+{}^RB_j')\right] | i,j \qquad \text{A2.6}$$

For the covariance expressions involving balancing areas b and c, V_s is replaced by Cov_s, $\sum_{i=1}^{N_b}$ does not exist and $\sum_{i\neq j}^{N_b}$ is replaced by $\sum_{i=1}^{N_b} \sum_{j=1}^{N_c}$.

$$\text{Cov} \left\{ n_b , \quad \sum_{i=1}^{n_b} \Pi_i^{-1} \left[\delta_i(X_i+{}^RB_i)+\delta_i''(X_i+{}^RB_i')\right]\right\}$$

$$= \text{Cov}_s \left\{ n_b , \quad \sum_{i=1}^{n_b} \Pi_i^{-1} \left[\alpha_i(X_i+{}^RB_i)+\alpha_i''(X_i+{}^RB_i')\right]\right\} \qquad \text{A2.7}$$

$$\text{and finally, Cov } n_b , \quad \sum_{i=1}^{n_b} (\delta_i+\delta_i'') = \text{Cov}_s \left[n_b , \quad \sum_{i=1}^{n_b} (\alpha_i+\alpha_i'')\right] \qquad \text{A2.8}$$

Now $V(n_b y_b / z_b)$ may be written approximately as:

$$\frac{(En_b)^2 (Ey_b)^2}{(Ez_b)^2} \; V \left(\frac{n_b}{En_b} + \frac{y_b}{Ey_b} - \frac{z_b}{E_b} \right)$$

Likewise, $Cov(n_b y_b / z_b, \; n_c y_c / z_c)$ can be approximately written as:

$$\frac{En_b}{Ez_b} \frac{Ey_b}{} \quad \frac{En_c Ey_c}{Ez_c} \; Cov \left(\frac{n_b}{En_b} + \frac{y_b}{Ey_b} - \frac{z_b}{Ez_b} \right), \left(\frac{n_c}{En_c} + \frac{y_c}{Ey_c} - \frac{z_c}{Ez_c} \right)$$

and by partial substitution of the formulae derived we may obtain $V(\hat{X}_b)$ as stated in 6.1 and $Cov \; (\hat{X}_b, \hat{X}_c)$.

For $V(\hat{X}_b)$ under methods (a) to (d), one simply puts all δ_i'', α_i'', $\bar{\alpha}_b''$ equal to zero. All $Cov \; \delta_i \delta_i''$, $Cov \; \delta_i \delta_j'' = 0$ for methods (a) to (d).