FIELD SUBSTITUTIONS - A NEGLECTED OPTION?

Vasja Vehovar, University of Ljubljana
Faculty of Social Sciences, 61109 Ljubljana, SLOVENIA

Key Words: Non-response, Missing Data

1 Introduction

Field substitutions occur when a non-responding unit is replaced with a substitute (reserve) unit during the field-work stage of the survey process. It seems that the practice of substitutions has already been superseded in survey research:

- textbooks generally do not recommend this procedure;
- often, substitutions are strictly forbidden in face-to-face surveys, especially if area frames are employed in probability samples.

However, on the other hand, we can observe the following:

- there is a relatively wide-spread use of substitutions in many important surveys in different countries;
- substitutions are particularly attractive in the case of CATI telephone interviewing, where the key drawback - problems with field-work controls - diminishes greatly.

In this paper we question whether substitutions can be justified, and if so, in what circumstances. We start with an outline of the problem (1). Next, substitutions are considered in the context of the unit non-response (2), the results from empirical evaluations with the Slovene national surveys are presented (3) and the bias-variance properties are discussed (4). At the end the conclusions (5) are summarized.

2 Substitutions, non-response

Unit non-response is commonly accepted term for an eligible unit which has been selected in a sample but which becomes missing in the field-work stage of the survey. This can introduce severe distortions into statistical inference, so a variety of ways of minimizing the problem have been developed in the design, field and processing stage of the survey process. Most commonly, some sort of "office" adjustments (direct methods, imputation, weighting) are used to compensate for missing data.

The substitution procedure is a very specific tool for coping with the unit non-response and, strictly speaking, it is a form of imputation.

There has been little written about this topic. Textbooks about survey methodology - and survey sampling in particular - mention substitutions very briefly - e.g. Kish (1965), Lessler (1991) - or not at all, e.g. Cochran (1978), Groves (1989). Generally, the literature doesn't recommend this option, at least in probability samples. However, it is mostly from the circumstances of area frames and face-to-face surveys that this opinion derives. With CATI telephone surveys or with sampling from population registers, the situation is slightly different.

An extensive search of the literature adds little to Chapman's research (1983) which is the most exhaustive treatment of the problem. Other important research was done at the US Census Bureau (Biemer, 1990), where substitutions were evaluated in the context of telephone survey methodology. However, the results were relatively ambiguous.

Other than this, only a few additional references can be found, e.g. Nathan (1980), Marliani (1993), Vehovar (1993), Forsman (1992). From a more practical point of view, the European Community's *Labour Force Survey* and the *Family Budget Survey* are extremely interesting. There, the practice of substitutions is observed in some European countries. Conditionally - under specific circumstances - substitutions are even recommended in the *Family Budget Survey* when the non-response rate exceeds 35% (Verma, 1993: 92).

3 Properties

The following properties of substitutions are partially derived from the available literature; the main sources, however, were extensive experiments with the Slovene national surveys - Labour Force Survey (LFS), Family Budget Survey (FBS), General Social Survey (GSS), Crime Victimization Survey (CVS) - during the years 1990-1993.

3.1 Advantages

First, let us briefly review the advantages:

Simplicity. When there are no other reasons for weighting, the substitutions provide a self-weighted sample and the merits of such samples are well-known. However, the avoidance of other non-response adjustments is very deceptive, since it is an empirical fact that substitutions cannot solve the entire non-response problem.

Sample size controls. Many authors emphasize the advantage of controlling the sample size and the interviewer's burden. This advantage is relatively minor since - in general - there also exists a variation in sample size arising from non-eligible units that cannot be controlled by substitutions. Furthermore, non-response still creates some uncontrolled differences in the number of units (visits) assigned to each interviewer. Also, there are severe practical difficulties in obtaining an exact sample size by means of substitution, even for telephone surveys.

Removal of the non-response bias. The available literature consistently claims that with substitutions we end up with respondents, so the non-response bias remains untouched. Often, however, this is not correct since there does exist an important improvement similar to that made with sample weighting adjustments. For example, the urban-rural component of the bias is generally removed when substitutions are made at the cluster level. Thus, it is only the non-response bias within the level at which the replacements are performed that remains untouched by substitutions.

In principle, we can further improve the non-response bias by selecting a substitute unit *similar* to the non-responding one. However, the shortage of available covariates, their weak correlation with the non-response characteristics and their weak correlation with the target variables make the actual non-response improvement much less successful. When combined with practical inconveniences in the selection of the *similar* substitutes, the discouraging results reported in **all** available empirical studies are not surprising.

Optimal structure of the sample. The substitutions provide the prescribed number of observations from each part of the sample. If the sample design is optimal, the optimality is also preserved in the obtained data. The non-response thus causes no distortion in the designed structure of the sample. This is, of course, irrelevant in the case of a simple random sample, but in complex designs this is an issue, especially when small strata or small clusters are employed at the last stage of selection and the non-response rate is high. In such situations, the substitutions might reduce the sampling variance.

The first three practical advantages above are generally not very significant, although situations may exist where they can be beneficial. Thus the issue of the optimal structure stands as the key advantage of the procedure.

Beside the above advantages, it appears that with substitutions we avoid completely the problem of having no observations (due to non-response) in certain areas, clusters or strata.

3.2 Disadvantages

There do, of course, exist severe disadvantages:

Field-work controls. This is the major deficiency of the substitutions, especially when area frames are used in face-to-face surveys. With telephone interviewing, however, the bulk of the problem disappears, as modern CATI systems provide excellent controls over data collection procedures.

The illusion that a non-response problem has been solved. The illusion that by substitutions we have solved the non-response problem is extremely strong. As a consequence, the effort to handle non-response may be reduced, or may not be made at all.

Higher non-response rate. An interviewer's effort decreases if he knows that difficult-to-contact units can be declared non-interviews and then replaced by substitutions. Of course, the fact that substitutions occur more often in areas with high non-response should be carefully taken into account.

Prolongation of the field work. According to the number of waves of substitutions, the field work is substantially prolonged. Conservatively speaking, each wave can be treated as a separate survey with the same number of prescribed attempts (visits, calls). So with substitutions, the field-work time may equal that needed for two or even three surveys.

Early respondent effect. In later waves of substitutions, the interviewers may select easy-to-contact units more often than in the initial sample. Typically, with substitutions the number of call-backs is smaller or no attempt to convert refusals is made.

Thus, the easy-to-contact units penetrate the survey to a greater extent than in the samples without substitutions. When the easy-to-contact units differ, there exists a risk of additional bias - a **net substitution bias**.

3.3 Practical guide-lines

We can summarize the following practical guide-lines:

1. Field substitutions are **not** appropriate for (large) probability samples where at least one of the following features is valid:

    - there is a short time available for field operations with a definite time-limit,
    - the early respondent effect exists,
    - there is weak control over field-work procedures (e.g. area frame).

2. The following practical reasons can justify the use of substitutions:

    - the need for a self-weighted sample is extremely strong; however, in this case the following conditions must additionally hold for the substitution procedure to possibly have some practical advantages:
        - there are no other theoretical reasons for weighting,
        - there is no serious non-response bias, or, if there is, the substitutions can remove this bias.
    - a danger exists that, because of non-response, a considerable number of strata or clusters would have no observations.

3. If we consider the improved precision arising from the optimal structure of the sample, the substitutions do have a theoretical advantage, however, its practical importance is a complex issue depending on the similarity of units within the last stage clusters (intracluster correlation), the size of a *take* per cluster, the level of the response rate and the level of the *net substitution bias*.

While the first two conclusions above (1, 2) are of a practical nature and have already been discussed, the last (3) needs some further consideration.

4 Bias and variance

With substitutions, the designed structure of the sample is actually obtained, and this should be manifested in the lower sampling variance and/or lower mean squared error (MSE) compared to alternative procedures.

In the following discussion, we analyze the unconditional mean and variance over repeating samples and over repeating non-response deletion. We restrict ourselves to the estimates of the population average. Basically, the estimate based on substitutions will be compared with the estimate based on the **corresponding** sample of respondents. By "corresponding sample of respondents" we understand that the initial sample size is $n^* = n/\bar{R}$, where $\bar{R}$ stands for response rate and n for sample size obtained with substitutions.

We will treat the bias in the general case, however, with variance we accept as plausible the assumption that data are *missing completely at random* (MCAR) within the level (area, cluster, strata) where the substitutions are performed (Little &, Rubin, 1987). Such an assumption enables us to incorporate the missing data mechanism (unit non-response) into inclusion probabilities of the *corresponding* sample of respondents. In practice, this is a common assumption leading to sample non-response weights attached to the respondents.

For simplicity, we also assume that the substitute units have the same interviewing costs as the initial units.

4.1 Simple random sample (SRS)

First, let us formalize the notion of *early respondent effect*. We split the initial respondents into secondary respondents and secondary non-respondents according to their behavior when contacted as substitute units. Thus, secondary non-respondents would respond if included in the initial sample, but they would not respond if selected as substitutes. Contrary to this, the secondary respondents would respond on both occasions. Then, instead of the well-known expression (Cochran, 1978: 359) for the *non-response bias*:

$$Bias(\bar{y}_{NON}) = \bar{M}(\bar{Y}_r - \bar{Y}_n),$$

where $\bar{M} = (1 - \bar{R})$ is the non-response rate, $\bar{Y}_r$ and $\bar{Y}_n$ denote the population average for respondents and for non-respondents, we have the gross *substitution bias*. It can be shown that it is of the following form:

$$Bias(\bar{y}_{SUB}) = \bar{M}(\bar{Y}_r - \bar{Y}_n) + \bar{M}\bar{M}_{sn}(\bar{Y}_{sr} - \bar{Y}_{sn}), \quad (1)$$

where $\bar{M}_{sn}$ stands for the proportion of secondary non-respondents among all initial respondents, and $\bar{Y}_{sr}$ and $\bar{Y}_{sn}$ denote the population averages for secondary respondents and secondary non-respondents.

As the $\bar{Y}_{sn}$ generally lies between $\bar{Y}_{sr}$ and $\bar{Y}_n$, the substitution bias $Bias(\bar{y}_{SUB})$ exceeds the non-response bias $Bias(\bar{y}_{NON})$. Due to the product $\bar{M}\bar{M}_{sn}$, the **net** ***substitution bias*** $\bar{M}\bar{M}_{sn}(\bar{Y}_{sr} - \bar{Y}_{sn})$ will be small and will be dominated by sampling error, at least in surveys up to a few thousands units.

It was found that, in the highly controlled face-to-face surveys with sampling from the register, the net substitution bias was around 0.5% of the estimate for the variables sex, age (group 18-25), rural-urban component, education.

The expression for the (unconditional) variance of the estimate $\bar{y}_{SUB}$ is relatively complex. It will not be discussed here, since the differences compared to the *corresponding* sample of respondents are considerably smaller than in the case of the bias. They are thus practically negligible.

4.2 Two-stage sample

Most often, the substitutions are performed at the certain (area) cluster level. There, the most appealing alternative procedure is a sample weighting adjustment at the same level with the weights inversely proportional to the estimated response rate in the cluster.

The non-response bias arising from the different response rates across clusters will be removed by substitutions, as with sample weighting adjustment. With respect to the remaining within-cluster component of the non-response bias, the results from the SRS sample apply (1).

Let us concentrate on variance. For simplicity we assume no non-response bias within clusters. Also, we will disregard the differences between the sampling variance based on substitutions and the case where there is no non-response; it can be shown that they are practically negligible.

With given inclusion probabilities at the first stage of selection the common estimators (Cochran, 1978, chp. 10, 11) of the population average have the following sampling variance structure:

$$Var(\bar{y}) = \frac{1}{a}U + \frac{1}{a}\sum_{i=1}^{A} V_i\left(1 - \frac{m_i}{M_i}\right)\frac{1}{m_i}, \qquad (2)$$

where a and A are numbers of clusters in the sample and in the population, U and V_i are quantities depending only on the population values, and m_i is the **designed** sample size within cluster. Of course with substitutions, we obtain the whole of the designed size m_i. Without substitution, but with the *corresponding* sample of respondents (initial size n^*), we have m^*_{ri} respondents with $E(m^*_{ri}) = m_i$.

The first term in (2) is the same for both procedures; however, unless the substitutions compensate for non-response the second term varies with the *take* m^*_i per cluster. Since the estimator $\bar{y}$ is unbiased, only the expected value of the expression (2) needs to be considered for the variance of the weighted sample.

It can be shown that, in the simple but realistic case where V_i and m^*_{ri} are independent, the factor $(1 - \frac{m^*_i}{M_i})$ is negligible or constant, and the non-response mechanism is a uniform Bernoulli mechanism with parameter $\bar{R}$, the increase in variance is based on:

$$E\left(\frac{1}{m^*_{ri}}\right) \Big/ \left(\frac{1}{m_i}\right). \qquad (3)$$

In Table 1, the increase (3) in the variance of the *corresponding* sample of respondents over the variance based on substitutions is presented. The calculations are based on a truncated hypergeometric distribution for the simplest self-weighted case, $m^*_i = \bar{m}^*$, with population cluster size $M_i = \bar{M} = 1000$. In the case of $\bar{M} = 100$, the figures in Table 1 would be roughly 10% lower. For large M_i, approximation with binomial distribution and Taylor linearization can be used (Cochran, 1978: 135).

	$\bar{R}$ - response rate				
$\bar{m}^*$	0.9	0.8	0.7	0.6	0.5
3	5.7	11.2	(16.0)	(18.2)	(18.2)
4	3.9	9.2	14.9	(19.4)	(21.7)
5	2.8	6.9	12.3	18.0	(22.4)
10	1.3	2.6	4.7	8.0	13.0
15	0.7	1.6	2.7	4.4	7.1
30	0.3	0.6	1.1	1.7	2.5

Table 1: Increase in variance (%) at $M_i = 1000$

The brackets in Table 1 indicate that more than 1% of the clusters were omitted (truncation) because no unit in the cluster responded.

Of course, the above increase refers only to the second component in (2), i.e. the within variance component. Its proportion can be expressed as a function of the intracluster correlation ρ and the size of the cluster. In the case of equal clusters, constant sampling rates within clusters and sampling without replacement at both stages, we can use the well-known relations from Kish (1965) to obtain results in Table

2. General experience with the target variables in official surveys shows that about a half of the sampling variance belongs to the within component.

$\bar{m}$	ρ - intracluster correlation					
	0.005	0.01	0.02	0.05	0.1	0.2
3	0.99	0.97	0.94	0.86	0.75	0.57
4	0.98	0.96	0.93	0.83	0.70	0.51
5	0.98	0.95	0.91	0.79	0.64	0.44
10	0.95	0.91	0.83	0.66	0.47	0.29
15	0.93	0.87	0.77	0.56	0.38	0.21
30	0.87	0.77	0.62	0.39	0.23	0.12

Table 2: Proportion of the within variance

As an approximation, the results from Table 1 and Table 2 can be used for other sampling strategies within the two-stage sampling scheme.

4.3 Discussion

We conclude that there may exist some special situations (i.e. small $\bar{m}$, small ρ, large $\bar{R}$) with a noticeable increase in variance, but in general the increase will be small. The substitutions thus have only a minor advantage in precision compared to the alternative sample weighting adjustment.

In the case of more than two stages, only primary sampling units (PSU) are important. However, with more stages, the PSU's will be generally large and the corresponding increase will be relatively small.

The above results were verified by six recent GSS surveys (n=2,100) where substitutions were used on a regular basis and the response rate was $\bar{R} = 0.8$. Thirty target variables were selected from each survey, and the variances were compared (weighting vs. substitutions). The effect of the different sample size was carefully removed using design effect and intracluster correlation. With clusters of $\bar{m} = 15$ the median increase was 2%, while with $\bar{m} = 5$ the median increase due to weighting was 10%. Compared to the corresponding ρ the increase was slightly above the theoretical results and some considerable variations around the median were observed. This is attributable to some other factors such as correlation between the weights and the variables.

The computer simulations (S-PLUS) of the GSS design confirmed the above conclusions much more clearly. However, an extremely large number of samples - at least 30,000 - was needed to obtain stable variance estimates.

A specific situation of considerable practical importance can be seen with the unweighted sample of respondents. There, the ratio estimator must be treated separately from the Horwitz-Thompson estimator which may suffer from a considerable increase in variance when compared with substitutions.

Also, it should be emphasized that with variable weights arising from sample weighting adjustment at the cluster level, the increase in variance may differ from the more common situation with fixed weights (w) based on oversampling strata. In that situation, the approximation:

$$VIF = 1 + relvar(w) = \sum w_k^2 \Big/ \left(\sum w_k\right)^2 \quad (4)$$

can be used (Kish, 1965).

Table 3 illustrates the increase in variance based on the above expression with the weights arising from the sample non-response adjustment in a two stage sample. We assume equal clusters, uniform sampling rates and uniform Bernoulli non-response mechanism. The weights are constructed at the cluster level proportionally to $1/m_{ri}^*$.

The figures from Table 3 can be compared with the increase based on Table 1 and Table 2. Since different principles are used, the figures also differ. It seems that the expression (4) considerably overestimates the increase in variance. However, in many practical situations the increase will be small in both situations, the differences thus being negligible.

$\bar{m}^*$	$\bar{R}$ - response rate				
	0.9	**0.8**	**0.7**	**0.6**	**0.5**
3	10.2	17.6	(19.4)	(17.6)	(16.0)
4	6.8	16.8	22.1	(24.0)	(22.0)
5	4.0	12.5	21.1	127.0	(26.5)
10	1.5	3.6	7.3	14.4	24.0
15	0.1	2.0	4.0	6.8	12.5
30	0.0	01.0	1.7	2.7	4.0

Table 3: Increase (%) in variance based on (VIF-1)

The issue of variable weights is rather complex, but it does not change our basic conclusions. In any case, the variable weight component should be treated separately from the fixed weight component arising from oversampling strata.

5 Conclusions

We conclude the following:

• With substitutions, the advantage of improved precision is relatively small since the increase in variance arising from the alternative sample weighting adjustment is generally very low. This finding contradicts the opinion that non-response adjustment at the level of small clusters substantially increases the variance. While there are some exceptions (small *take* per cluster, high non-response rate, small intracluster correlation), the increase will rarely exceed 10%.

• For large surveys ($n > 10,000$) a small but consistent *net substitution bias* will generally dominate over a negligible increase in precision which occurs in the case of corresponding weighting procedure. Thus, the optimization of the MSE doesn't favour substitutions in these situations.

• In surveys up to a few thousand units, with proper field-work controls, the bias-variance issues become unimportant in most situations for the evaluation of the substitutions. There, the practical considerations play the key role. However, when there is not enough care in the field-work procedures, we can easily end up with a considerable *net substitution bias*.

• The prolongation of the data collection period stands - together with the inconveniences in the field-work process - as the key practical drawback of substitutions. From the same practical viewpoint, substitutions may be, conditionally (section 3), advantageous when a strong need exists for the self-weighted sample.

• It seems that substitutions are really needed only in surveys with high non-response rates and small *take* per cluster (or strata) where we may have no observations. We encounter this situation in surveys of institutions, e.g. stores or schools, or in household surveys such as *Family Budget Surveys*.

• We have dealt with a simple method for dealing with non-response. Nevertheless, other adjustments, such as population weighting, may follow in both cases (i.e. substitutions and *corresponding* weighting adjustment). Of course, when models are used, such as those of Brehm (1993) or Little & Rubin (1987), substitutions are simply redundant.

References

[1] BIEMER P, Chapman D., Alexander C. (1990) *Some Research Issues in Random-Digit-Dialing Sampling and Estimation.* 1990 Annual Research Conference, Washington, Bureau of the Census.

[2] BREHM J. (1993) *The Phantom Respondents.* The University of Michigan Press, Ann Arbor.

[3] CHAPMAN D. (1983) *The Impact of Substitutions on Survey Estimates.* Meadow W.G., Olkin I., Rubin D., Eds.: Incomplete Data in Sample Surveys, Vol.II, Theory and Bibliographies. New York, National Academy of Sciences, Academic Press, pp. 45-61.

[4] COCHRAN W. (1978) *Sampling Techniques.* Wiley, New York.

[5] FORSMAN G., Berg S. (1992) *Telephone Interviewing and Data Quality, an Overview and Empirical Study.* Linkoping University.

[6] GROVES R. (1989) *Survey Errors and Survey Costs.* Wiley, New York.

[7] KALTON G. (1983) *Compensating for Missing Survey Data.* Institute for Social Research, Ann Arbor.

[8] KISH L. (1965) *Survey Sampling.* Wiley, New York.

[9] LESSLER J. (1991) *Non-sampling Errors in Survey.* Wiley, New York.

[10] LITTLE R., Rubin D. (1987) *Statistical Analysis with Missing Data.* Wiley, New York.

[11] MARLIANI G., Pacei S. (1993) *Effects of Household Substitution in the Italian Consumer Expenditure Survey.* Bulletin of the ISI, Contributed Paper, 49th ISI Session, Firenze 1993, p. 149-150.

[12] NATHAN G. (1980) *Substitution for Non-response as a Means to Control Sample Size.* Sankhya, C42 (1980), 1-2, p. 50-55.

[13] VEHOVAR V. (1993) *The Field Substitution in the Slovene Labour Force Survey.* Bulletin of the ISI, Contributed Paper, 49th ISI Session, Firenze 1993, p. 519-520.

[14] VERMA V., Gabilondo L.G. (1993) *Family Budget Surveys in the EC: Methodology and Recommendations for Harmonization.* Eurostat, Statistical Document 3 E, Luxembourg.