Contributions to Methodology and Statistics
A. Ferligoj and A. Kramberger (Editors)
Metodološki zvezki, 10, Ljubljana: FDV 1995

Field Substitutions in Slovene Public Opinion Survey

Vasja Vehovar[1]

Abstract

We talk about field substitutions when a non-responding unit is replaced with a substitute (reserve) unit in the field-work stage of the survey process. The case of the Slovene Public Opinion (SJM) survey serves as a starting point for bringing the issue of substitutions into the broader context of missing data problems. In empirical evaluation the substitution procedure is compared with the alternative weighting adjustment. The main finding is that the advantages of the substitution procedure can hardly compensate for its serious drawbacks. Specifically, the comparable gains in precision are relatively small, at least when a ratio estimator is used for the estimation of the population average. However, the specific setting of the Slovene Public Opinion survey makes the substitutions still a relatively favorable procedure. In any case, the substitution procedure is becoming more and more obsolete as it is relatively inefficient in handling the growing non-response problem.

1 Introduction

The Slovene Public Opinion (Slovensko Javno Mnenje - SJM) survey, administered at the Public Opinion and Mass Communication Research Center, Faculty of Social Sciences at University of Ljubljana, is the most important social survey project conducted in Slovenia. The methodology employed, which has remained unchanged for a quarter of a century, includes the use of field substitutions. Therefore, our initial question will be: *Can we justify the substitution procedure in the context of contemporary survey methodology*?

There is a lot of evidence that substitutions - as a solution to the non-response problem - have already become outmoded in survey practice. For example, textbooks generally do not recommend this practice. Also, in many countries - the USA is the best example - substitutions are practically "forbidden" in face-to-face

[1] Faculty of Social Sciences, University of Ljubljana, P.O. Box 47, 61109 Ljubljana, Slovenia

surveys (due to problems of field-work), especially when area frames are employed. On the other hand, there is a relatively wide-spread use in Europe and also in many other countries. Within this contradictory situation, we are looking for an answer to our question which we can apply to the SJM survey. Also, conclusions are sought for the more general case of probability sample surveys. Basically, however, we compare the substitution procedure with the simple weighting adjustment.

We start with a description of the substitution procedure in the SJM survey (1). Next, the substitutions are considered in the general context of missing data problems (2). The issue of bias and variance trade-offs is addressed in more detail with an empirical evaluation of the SJM survey (3). The substitutions are compared with the simple weighting procedure with the weights inversely proportional to the response rate within primary sampling units. After results (4) and discussion (5) the conclusions (6) are summarized.

2 Slovene Public Opinion survey

The Slovene Public Opinion (SJM) is a standardized face-to-face survey performed regularly a few times a year. Its core content is close to the "General Social Survey". However, it also resembles omnibus surveys in that it serves the needs of international survey projects and of public clients.

2.1 Sample design

The sample design (Blejec, 1970) is a three-stage probability sample with implicit regional stratification. Thus, there is no explicit stratification. The SJM survey started with the election lists as the sampling frame, but switched a few years ago to the central register of population. There, the name, address, age and gender of the selected persons are available. There are 140 primary sampling units[2] (PSU) selected in the SJM sample. Within each PSU, three secondary sampling units[3] (SSU) are selected. Five persons (c=5) are finally selected from these 140x3 SSU, thus giving a sample of 140x3x5=2,100 persons. Experience shows that the median of design effect is around deff ≈ 1.5.

[2] The PSUs are the clusters of 4,200 adjacent persons in the register of the Slovene population (age 18-72). The register is sorted by communities, settlements, streets and house numbers.

[3] The SSU are clusters of 100 adjacent persons within the PSUs.

2.2 Substitutions

Whenever an interview with the selected person from the sample cannot be obtained we have a missing unit. Let us label this a non-interview. There are a variety of reasons for a non-interview. A person may die, or a person may be non-eligible for some other reason (e.g., movers) - we then have a non-eligible unit. Also, the interview may not be obtained from the eligible unit due to non-contact, refusal or inability to cooperate - in which case we have a (unit) non-response. In addition, there may exist some other specific (but minor) reasons for a non-interview.

Let us consider a typical SJM survey - SJM 1991/2. The 2105 persons selected in the initial sample resulted in 1677 responding units and 428 non-interviews. The non-interviews can be further classified as 212 non-respondents, 181 non-eligible persons and 35 non-classified non-interviews. The non-respondents consist of 73 refusals, 113 non-contacts and 26 other reasons for non-response. We can calculate the non-response rate[4] of 11% and the refusal rate of 5%. The total completion[5] rate was 80%.

The reasons for a non-eligible unit can be further classified in four approximately equal subgroups:

- persons living in other countries, but formally residents (voters) of Slovenia,
- persons living in Slovenia at an actual address different from their formal address,
- non-accuracy (i.e. movers, deaths) of the register,
- residual cases (unknown, unclassified, nonexisting,...).

Of course, strictly from a sampling point of view, non-eligible persons must be ignored. If substitutions were used, the PSUs with a high percentage of non-eligible persons would obtain an inclusion probability which was too large. As already mentioned, in the SJM survey, however, substitutions are used for all non-interview cases and not only for non-response. For this reason, in addition to the initial sample of 5 units within each SSU, another 4 addresses are randomly selected as a potential substitute (reserve) units. Whenever a non-interview occurs,

[4] We assume that the non-classified non-interviews are distributed in similar manner to the classified ones. The non-response rate is calculated as the ratio of the non-respondents to the eligible units 212/(2105-35-181)=0.11.

[5] The completion rate is the ratio of the number of respondents to the number of all units included in the sample (1677/2105=0.8)

a substitute unit is used. To prevent "easy" replacements, up to five attempts (visits) have to be made[6].

Generally, two or even three waves of substitutions have to be made in the SJM survey in order to obtain five interviews from each SSU. Schematically, the situation is as follows:

Table 1A: The substitutions in the SJM 91/2 survey

	initial units	substitute units	total
respondents	1677	401	2078
non-interviews	428	146	574
total	2105	547	2652

We should add that out of 547 persons from the reserve sample 428 were selected as a first wave of substitutions (instead of initial 428 non-interviews) and 119 (=547-428) were selected as the later waves of substitutions.

3 Missing units and field substitutions

3.1 Missing unit problem

We label as a missing unit (non-interview) any unit included into the sample but not interviewed. As the problem of the non-eligible units has a simple solution[7], the key example of the missing unit remains a non-responding unit or unit non-response - an eligible unit that was missed in the field-work process. The non-responding units create severe difficulties for the process of statistical inference since we cannot draw proper conclusions about the target population when there is non-response. Also, the sample size becomes a variable quantity. When there is a considerable number of non-responding units and they differ significantly from the respondents, the standard procedures of statistical inference may lead to serious distortions in conclusions.

There are many approaches for handling the non-response problem. The broadest classification could be as follows: ignoring the problem, "office" adjustments, special survey-design techniques, active field-work strategies, use of field substitutions.

[6] A control letter was sent immediately after survey to all selected persons asking about the interviewer's visit. (A pre-letter was also sent to the selected persons a week before the survey.)

[7] As mentioned, the non-eligible units should, in principle, simply be ignored. Of course, the eligible units not in the frame are the other side of this problem, one which leads to specific frame and non-coverage issues. However, we are concerned here only with probability samples where all the units have known and positive inclusion probability.

The survey theory is predominantly concerned with the "office adjustments", i.e.: direct methods, imputation methods and weighting methods (Little & Rubin, 1989). The special survey design techniques include: double sampling for non-response, over-sampling of domains with high non-response, the Politz-Simons procedure, etc. (Madow et al., 1983). The active field-work strategies cover: role of the interviewer, incentives, call-back strategies, field-work controls, etc. (Report, 1993). Here, of course, we focus on the substitution procedure.

3.2 The issue of field substitutions

The problem of field substitutions can be viewed from many different angles. One aspect which is often over-exaggerated is the following: *Can the substitutions compensate for the non-response?* Frequently, field substitutions are judged by this very criterion. However, regardless of this aspect, there are many other important features of this procedure and the basic question should be: *Are field substitutions advantageous compared to alternative procedures?*

There is little written about substitutions. General textbooks about survey methodology - and survey sampling in particular - mention the substitutions only briefly - e.g., Kish (1965), Yates (1960), Kalton and Moser (1971), Foreman (1991), Lessler (1991) - or not at all, e.g., Cochran (1977), Sardnal et al. (1992), Groves (1989). The brief but exhaustive treatment in Kish (1965) has become, more or less, the standard attitude: due to the problems with field-work controls the substitutions are not a recommended option in probability samples. Sometimes, however, there is a promise of the removal of the non-response bias (Lessler, 1991: 177).

An extensive search for published work in this area adds little to Chapman's research (1983), which is the most exhaustive treatment of this problem. Other important research was done at the US Census Bureau (Chapman and Roman, 1985), where substitutions were evaluated in the context of telephone survey methodology. Besides this, only a few additional references can be found: Forsman (1992), Marliani & Pacei (1993), Nathan (1980). From a more practical point of view, the Eurostat's Labour Force Survey (Verma, 1992) and the Family Budget Survey (Verma, 1993) are extremely interesting. There, the practice of substitution is observed in two important European official surveys. Conditionally, substitutions are even recommended in the Family Budget Survey if the non-response rate exceeds 35% (Verma, 1993: 92).

3.3 Features of the substitutions

The following properties of the substitution procedure are partially derived from the available literature, especially Chapman (1983); the main sources, however, were the results of studies of the Labour Force Survey, Family Budget Survey and Slovene Public Opinion survey data during the four years from 1990-1993. Here is a brief review.

3.3.1 Advantages

a) Simplicity: when there is no other reason for weighting, the substitutions provide a self-weighted sample which is free from the difficulties imposed by weights.
b) Sample size controls: some authors (Chapman, 1983) emphasize the advantage of controlling the sample size and the interviewer's burden.
c) Removal of the non-response bias: if we select a substitute unit similar to the non-responding one then the non-response bias may be reduced.
d) Optimal structure of the sample: if we design the optimal sample, the optimality is preserved in the obtained data, too - the non-response thus causes no distortion in the designed structure of the sample.

The first three advantages above (a-c) are generally not very persuasive, although situations may exist where they can be beneficial. Thus, the advantage of the optimal structure stands as the key advantage of the substitution procedure.

3.3.2 Disadvantages

a) Field-work controls: the interviewers are always left with a relatively high and hard-to-control responsibility, at least in face-to-face surveys.
b) The illusion that a non-response problem has been solved: the effort to handle the non-response problem might be less intensive or may even diminish completely.
c) Higher non-response rate: the intensity of the interviewer's effort decreases if the difficult-to-contact units can be declared as non-interviews and then replaced by substitutions.
d) Early respondent effect: there may exist an additional bias arising from the substitution procedure since in the later waves of substitutions the interviewer may select the easy-to-contact units more often than in the initial sample. Thus, the easy-to-contact units penetrate the survey to a larger extent compared to a situation without substitutions.

3.3.3 General conclusions about field substitutions

Here are the most important findings concerning the practice of the substitutions:

a) The use of field substitutions is not appropriate for (large) probability samples with at least one of the following features valid:
  i) there is a relatively short time available for field operations with a definite time-limit,
  ii) there exists the early respondent effect,
  iii) there is weak control over the field-work procedures.

b) The following conditions must additionally hold for the substitution procedure to possibly have some practical advantages:
  i) there are many "simple" users of the sample survey data,
  ii) there are no other theoretical reasons for weighting,
  iii) there is no non-response bias or, if there is, the substitutions can remove this bias.

c) If we favour substitutions because they preserve the designed structure of the sample, the following features become important:
  i) the similarity of units within the last stage clusters (intra-cluster correlation),
  ii) the size of a "take" per last stage cluster,
  iii) the level of the response rate.

While the first two properties (a,b) relate to more general methodological issues and have been discussed briefly in the previous section, the last paragraph (c) relates to the statistical properties of the substitution procedure. In the next section these properties will be discussed. However, we will restrict the discussion to the empirical evaluation of the SJM survey.

4 Methodology

As mentioned, the key advantage of the substitution procedure is the fact that it maintains the designed structure of the sample. Of course this should be manifested in the lower mean squared error (MSE) compared to the alternative procedures. Here, we will use the SJM survey to compare the substitution procedure and the weighting adjustment which is the recommended adjustment in the case of unit non-response (Kalton, 1983). For the purpose of empirical evaluation the following SJM surveys were used: SJM 1990/1 (n=2045), SJM 1990/2 (n=2072), SJM 1991/2 (n=2072), SJM 1992/1 (n=2084), SJM 1992/2 (n=2023), SJM 1993/1 (n=2058). About 30 target variables (proportions) were selected from each survey. We considered more than one survey to ensure stable conclusions. For each variable the estimates based on the sample with substitutes were compared (bias, variance) with the estimates based on weighting adjustments.

Since the sample was drawn from the national register of population (N=2,000,000), the true individual data were available for the following three variables: gender, age and type of settlement (rurality). Thus, the above three control variables were also compared with their population values.

The detailed analysis of non-interview and non-response in the SJM survey revealed certain differences between non-contacts and refusals. However, the problem of non-response was not the main focus of this study. Here, we were predominantly concerned with the comparison of the two simple and practical procedures (substitutions, weighting) as they are commonly used in practice.

Similarly, we treated the field-work strategy as given and we did not discuss the possible cost-effects of the increased number of contacts or the effects of the refusal conversion strategy.

We should add that we assume that the costs are the same for both procedures. There is no evidence in the SJM survey that substitutions are more costly than weighting. To achieve a certain number of responses, the expected number of contacts needed is the same for both procedures. However, with substitutions, higher overhead costs may arise from the complexity of the procedure or from the prolonged data collection stage.

4.1 Bias

Here we are dealing only with proportions. We assume a two-stage cluster sample of size *n*. We define a non-response bias as a difference between the estimate p_{RESP} based on the responding units and the estimate p_{TOT} based on the whole sample (respondents and non-respondents):

$$\text{Bias}_{NON}(p) = E_1(p_{TOT}) - E_1(E_2(p_{RESP})).$$

The expectation E_1 is calculated across all possible realizations of a specific sample and the expectation E_2 is calculated across all possible realizations of a particular missing data mechanism. For simplicity we will treat a sample size *n* and number of respondents *m* as fixed. In instances where the adjustments are made for all non-interview units it is more proper to label $\text{Bias}_{NON}(p)$ as a non-interview bias. We can separate two components of the non-response bias:

$$\text{Bias}_{NON}(p) = \text{Bias}^{(1)}_{NON}(p) + \text{Bias}^{(2)}_{NON}(p).$$

$Bias^{(1)}_{NON}(p)$ is the bias due to the differences between the non-respondents and respondents within clusters and $Bias^{(2)}_{NON}(p)$ is the bias which arises because the response rates differ across the clusters. The full elaboration of both components can be found in (Falorsi & Russo, 1993). We can thus write:

$$P = E_1(p_{TOT}) = E_1(E_2(p_{RESP})) + Bias^{(1)}_{NON}(p) + Bias^{(2)}_{NON}(p).$$

By definition, the weighting adjustment with the inverse of the response rate at cluster level (direct estimator) removes the second component of the non-response bias:

$$E_1(E_2(p_{WGH})) = E_1(E_2(p_{RESP})) + Bias^{(2)}_{NON}(p).$$

In principle, the substitution procedure should also remove the second component of the non-response bias. We define a total of the substitution bias as:

$$Bias^{(T)}_{SUB}(p) = E_1(p_{TOT}) - E_1(E_2(E_3(p_{SUB})),$$

where p_{SUB} denotes the estimate based on the sample with substitutions and E_3 denotes the expectation across all possible substitutions (replacements) for missing units with a given sample of respondents. Of course, the substitution bias $Bias^{(T)}_{SUB}(p)$ also includes the first component of the non-response bias $Bias^{(1)}_{NON}(p)$ since substitutions - similar to the cluster weighting - remove only $Bias^{(2)}_{NON}(p)$. To separate $Bias^{(1)}_{NON}(p)$ from the substitution bias we introduce a net substitution bias $Bias^{(N)}_{SUB}(p)$:

$$Bias^{(T)}_{SUB}(p) = Bias^{(N)}_{SUB}(p) + Bias^{(1)}_{NON}(p).$$

We can describe this bias as an inability of substitutions to remove the component $Bias^{(2)}_{NON}(p)$ which is arising from the different response rates across clusters. This may happen because of the early respondent effects or because of some other distortions in the selection of the substitutes. In the following discussion the term ***substitution bias*** will refer only to the ***net*** substitution bias. It is useful to observe the biases in the context of the following equation:

$$P=E_1(p_{TOT})=E_1(E_2(p_{RESP}))+(Bias^{(2)}_{NON}(p) - Bias^{(N)}_{SUB}(p))+Bias^{(N)}_{SUB}(p)+Bias^{(1)}_{NON}(p).$$

If we assume that the units are missing completely at random (MCAR) within the clusters - which is equal to the assumption $Bias^{(1)}_{NON}(p)=0$ - we can use the weighted estimate p_{WGH} as an unbiased estimate of p_{TOT}. In other words: if we assume that units within the clusters are missing completely at random, then the weighting with the inverse of the response rate within clusters leads to the proper Horwitz-Thompsen estimator (Little and Rubin, 1987: 55). Of course, if the MCAR assumption does not hold, then neither weighting nor substitutions is the optimal approach for this type of non-response. However, the substitution bias $Bias^{(N)}_{SUB}(p)$ still measures the extra bias introduced by substitutions when dealing with the second component of the non-response bias. Of course, the substitutions cannot remove the first component of the non-response bias unless we have a controlled selection of the substitutes within clusters (for example, we replace a male non-respondent with another male from the same cluster).

To repeat, there is a difference when the weighting is performed for the non-interview cases compared to the weighting only for non-response. The difference will always be made explicit in our further discussion.

Within the weighting adjustment we can also estimate the proportion p_{NON-W} among the units that were implicitly imputed instead of the non-responding (or non-interviewing) units:

$$p_{WGH-I} = (np_{WGH} - mp_{RESP})/(n-m),$$

where n denotes the total number of units and $m=n_{RESP}$ the total number of respondents. Similarly, the substitute units can - and should - be treated as a special type of imputation procedure. We will denote the corresponding proportion among the substitute units as p_{SUB} and the proportion among initial non-interview units as p_{NON}. We will use the following expressions to estimate the biases:

$$\begin{aligned}
bias^{(2)}_{NON}(p) &= p_{WGH} - p_{RESP}, \\
bias^{(N)}_{SUB}(p) &= p_{WGH} - p_{SUB}, \\
bias_{NON}(p) &= p_{TOT} - p_{RESP}, \\
bias^{(1)}_{NON}(p) &= p_{TOT} - p_{WGH}, \\
bias^{(T)}_{SUB}(p) &= p_{TOT} - p_{SUB}.
\end{aligned}$$

Of course, the following equations must hold for the estimated quantities:

$$\begin{aligned}
bias_{SUB}{}^{(T)}(p) &= bias^{(N)}_{SUB}(p) + bias^{(1)}_{NON}(p), \\
bias_{NON}(p) &= bias^{(1)}_{NON}(p) + bias^{(2)}_{NON}(p), \\
p_{WGH} &= (mp_{RESP} + (n-m)p_{WHG-I})/n, \\
p_{TOT} &= (mp_{RESP} + (n-m)p_{NON})/n, \\
p_{SUB} &= (mp_{RESP} + (n-m)p_{SUB})/n.
\end{aligned}$$

Generally, we will express the biases in the relative terms:

$$\text{Rbias}(p) = \text{Bias}(p)/P,$$

with the corresponding estimate rbias(p) = bias(p)/p. The relative non-response bias has an appealing interpretation: it is the percentage of error due to non-response. Similarly, the substitution bias is the relative distortion introduced by substitution procedure.

4.2 Weights

The sample weighting adjustment at the cluster level is comparable with substitutions, because both procedures compensate for the missing data at the same level. Also, in both cases we assume that the units are missing completely at random (MCAR) at the cluster level (e.g., within PSU or SSU) where the weighting (or substitution) is performed. For example, in the whole sample the non-response may be correlated with the urban-rural characteristic of the units; however, within small and homogeneous clusters, where all the units are of the same rural-urban type, this link may disappear. Thus, the units would be missing MCAR within the clusters, although they are not missing MCAR in the sample.

We should repeat that in the SJM surveys the substitutions are selected within the SSU clusters, so the corresponding weighting adjustment should also be performed at the SSU level. However, for variance calculations only PSU level weights are important. Also, the SSU clusters within each PSU are close to each other, so the adjustment at the PSU level is justifiable as it may result in smaller variation of the weights.

The weights were calculated as an inverse of the response rate in each cluster. Assume, for example, that there are only 12 responding units among 15 initial units in a certain cluster. Thus, 3 replacements are needed in the case of substitution procedure. Similarly, in the case of weighting adjustment all 12 responding units will receive the weight proportional to 15/12. Of course, if the weighting is done only for the non-response, then we need to separate the 3 non-interview units into, say, one non-eligible unit and two non-responding units. In this case, the weights will be proportional to 14/12. The corresponding substitution procedure should also involve replacement of only two non-responding units. In case of the adjustments at the SSU level the same principle applies, the only difference being the fact that we have a cluster of 5 instead of 15 units. In the case of weighting adjustment within SSU clusters the variance was calculated assuming 420 PSU, each with 5 persons.

Throughout our analysis the weighting adjustments were made for all non-interview cases. Thus we will treat all non-interview cases as non-respondents. The most obvious justification for this is the simple fact of comparison: the substitutions were also taken for all non-interview cases. Further justification - including the empirical comparison with proper non-response adjustments - will be presented at the end of this section, where both the substitutions and the weighting will be considered for the non-responding cases only.

4.3 Variance

The variance of the ratio estimator was calculated with a standard variance estimation procedure (Fuller, 1989) where the weights were incorporated. When evaluating the increase in variance due to weighting we refer to the theory stating that with simple random sampling, the increase in the variance due to weights from over-sampling the strata (Kish, 1965: 427) equals:

$$VIF = \Sigma w_i^2/(\Sigma w_i)^2,$$

where w_i denotes a weight for the i-th unit. In practice, the above expression is widely used in many complex designs. Some variance inflation factors (VIF) based on the above equation can be found in Table 1. There, we assume a two-stage cluster design with a Bernoulli missing data mechanism within each cluster: the weights are constructed as described in the previous section. The increase in the variance thus depends on the cluster size and the response rate.

Table 1: Variance inflation factors (%) due to non-response in a cluster sample

Cluster	Nonresponse rate				
size	10	20	30	40	50
3	10.2	17.6	19.4	17.6	16.0
4	6.8	16.0	22.1	24.0	22.0
5	4.0	12.5	21.1	27.0	27.0
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	1.0	1.7	2.7	4.0

Of course, the above approximation may be seriously distorted because of the correlation between the weights and the key variables. Also optimal stratification effects of the weights or the optimal cluster allocation effects may occur. Thus, we sometimes obtain rather different and unpredictable results (Judkins, 1991: 586), although, as a general principle, the above approximation is reported to be robust (Kish, 1992).

4.4 Coefficient of variation

To evaluate the extent of the relative biases we have to compare them with the relative sampling error - coefficient of variation:

$$CV(p)=SE(p)/p,$$

with the estimate $cv(p) = se(p)/p$. We will often refer to the design effect:

$$deff(p) = var(p)/var_{SRS}(p) \quad \text{and} \quad deff(p) = 1 + roh(p)(b\text{-}1),$$

where *roh(p)* measures the intra-cluster correlation, the average correlation of the units within clusters and *b* denotes the size of the PSU. However, the coefficient of variation CV(p) defined above is not the coefficient of variation of the bias. To calculate the sampling error of the biases we use the following approximation:

$$\begin{aligned}
&var(bias^{(T)}_{SUB}) = var(p_{TOT} - p_{SUB}) = \\
&= var(\,(mp_{RESP} + (n\text{-}m)p_{NON})/n) - (mp_{RESP} + (n\text{-}m)p_{SUB\text{-}I})/n = \\
&= var((n\text{-}m)(p_{NON} - p_{SUB\text{-}I})/n) = (n\text{-}m)^2/n^2 var(p_{NON}\text{-}p_{SUB\text{-}I}) = \\
&= (n\text{-}m)^2/n^2(var(p_{NON}) + var(p_{SUB\text{-}I}) - 2cov(p_{NON},p_{SUB\text{-}I})).
\end{aligned}$$

Since we can assume the same sampling variability for the nonresponding and substitute units we have: $var(p_{NON}) \approx var(p_{SUB\text{-}I})$. We observe that there is only the correlation within independent clusters, so we have an intracluster correlation roh(p) and the covariance can be written: $cov(p_{NON},p_{SUB\text{-}I}) \approx roh(p)var(p_{NON})$. We obtain:

$$\begin{aligned}
&var(bias(T)SUB) \approx \\
&\approx 2(n\text{-}m)^2/n^2 var(pNON)(1\text{-}roh(p)) \approx \\
&\approx 2(n\text{-}m)^2/n^2(1\text{-}roh(p))(S^2(p)/(n\text{-}m)deffNON(\,p))) \approx \\
&\approx 2(n\text{-}m)/n^2S^2(p)(1\text{-}roh(p))(1+roh(p)(b(n\text{-}m)/n\text{-}1)).
\end{aligned}$$

The above expression also holds for the other biases. It is a conservative estimate: in practice there might be much less variability in the p_{NON}, p_{SUB}, p_{WGH}, because there exist additional restrictions in the missing data mechanism and in the substitution selection procedure. Comparing the above variance with $var(p)=S^2(p)/n(1+roh(p)(b\text{-}1))$ we obtain:

$$\begin{aligned}
F^2&=var(p)/var(bias^{(T)}_{SUB}(p))= \\
&= (1+roh(p)(b\text{-}1))/2(n\text{-}m)/n(1\text{-}roh(p))(1+roh(p)(b(n\text{-}m)/n\text{-}1)).
\end{aligned}$$

We thus have: $se(bias^{(T)}_{SUB}(p)) \equiv se'(p) = se(p)/F$, $cv'(p)=se'(p)/p$.
In the case of the SJM survey we have ($m/n=0.8$) and the following factors F:

- if $deff(p)=1.0$, $roh(p) = 0.00 \Rightarrow$
  $\Rightarrow F=1.6$, $cv'(p) = cv(bias^{(T)}_{SUB}(p)) \approx cv(p)/1.6$,
- if $deff(p)=1.7$, $roh(p) = 0.05 \Rightarrow$
  $\Rightarrow F=2.0$, $cv'(p) = cv(bias^{(T)}_{SUB}(p)) \approx cv(p)/2.0$,
- if $deff(p)=2.4$, $roh(p) = 0.10 \Rightarrow$
  $\Rightarrow F=2.5$, $cv'(p) = cv(bias^{(T)}_{SUB}(p)) \approx cv(p)/2.5$,
- if $deff(p)=8.0$, $roh(p) = 0.50 \Rightarrow$
  $\Rightarrow F=5.0$, $cv'(p) = cv(bias^{(T)}_{SUB}(p)) \approx cv(p)/5.0$.

Obviously, the $cv'(p)=cv(bias^{(T)}_{SUB}(p))$ are much lower than the corresponding quantities cv(p). We will compare the relative biases with both cv'(p), to observe the significance of the relative bias, and cv(p) to evaluate the overall importance of the bias.

4.5 Mean squared error

To compare the weighting and the substitutions we should consider both the variance and the bias expressed in the mean squared error:

$$MSE(p) = Var(p) + Bias^2(p).$$

Both procedures should then be described in terms of the relative increase of the MSE(p):

$$RMSE(p) = MSE(p)/P^2 = CV^2(p) + Rbias^2(p).$$

Of course, by the very definition, the weighting adjustment causes no increase in the bias component. According to the previous section the increase of the variance is measured as a variance inflation factor VIF, so we have:

$$RMSE_{WGH}(p) = CV^2(p)(1+VIF(p)).$$

Similarly, in the case of substitutions we have increase only in the component of the bias, since by definition VIF=0. We can write:

$$RMSE_{SUB}(p) = CV^2(p)(1 + Rbias^2(p)/CV^2(p)).$$

5 Results

5.1 Bias in control variables

In the SJM survey 91/2, the data from the register (gender, age and size of the settlement) were attached. For the other surveys only survey data were analyzed.

We can observe in Table 2 some key results regarding the bias of the control variables. Let us repeat that in Table 2, both the substitutions and the weighting adjustment were performed for all non-interview cases. The weighting adjustments in Table 2 were done at the PSU level. The weighting adjustments at the SSU level were also performed; however there were no differences in terms of bias. There were, of course, considerable differences in variance, and these will be discussed in the next section.

Table 2: Control variables in the SJM 91/2 survey (%)

variable	initial sample	initial respondents	non-interview	implicit imputation by weighting	substitution (respondents)	FINAL SAMPLE substitutions	FINAL SAMPLE weighting
notation	p_{TOT}	p_{RESP}	p_{NON}	p_{WGH-I}	p_{SUB-I}	p_{SUB}	p_{WGH}
n	2105	1677	400[8]	400	400	2077	1677
Gender (F)	50.9	53.0	41.7	53.5	51.3	52.7	53.1
Age 18-30	26.5	26.3	27.7	25.8	25.8	26.2	26.2
Rural (<2000)	52.1	53.4	47.7	47.3	49.5	52.8	52.3
Education (>12)	-	34.5	-	34.6	38.5	36.0	35.8

5.1.1 Gender

It can be observed from Table 2 that respondents are more likely to be female. The differences - consistent across all SJM surveys - are, in fact, dramatic: the percentage of females among the initial respondents is p_{RESP}=53.0%, but only p_{NON}=41.7% among non-interviews. Also, we can observe that among the (responding) substitute units the bias towards women is slightly less intensive - we have $p_{SUB-I}(p)$=51.3%. However, we can observe from Table 4 - which shows the results for all 6 SJM surveys - that SJM 91/2 is an exception. The proportion of females among the substitute units is, in fact, generally higher than among respondents. In any case, substitutions failed to remove the non-interview bias. The same is true for the weighting procedure: the implicit imputation for non-interview cases gives a percentage of p_{WGH-I}=53.5%. The final estimate with substitutions is p_{SUB}=52.7%, and the weighted estimate is p_{WGH}=53.1%. The weighting generally performs slightly better, the SJM 91/2 being the only

[8] Only first wave non-interviews are included here. Similarly, with substitution, only the responding substitute units are presented.

exception. However, the true value of the initial sample is $p_{TOT}=50.9\%$. The population value, of minor importance here, is $P=51.5\%$.

Table 3: Absolute (relative) biases in the control variables for the SJM 91/2 survey (%)

variable	$bias_{NON}(p)$	$bias^{(1)}_{NON}$	$bias^{(2)}_{NON}$	$bias^{(T)}_{SUB}$	$bias^{(N)}_{SUB}$
Gender (F)	-2.1 (4.1)	-2.2 (4.3)	0.1 (0.2)	-1.8 (3.5)	0.4 (0.7)
Age 18-30	0.2 (0.9)	0.3 (1.0)	-0.1 (0.4)	0.3 (1.1)	0.0 (0.0)
Rural (<2000)	-1.3 (2.5)	-0.2 (0.4)	-1.1 (2.1)	-0.7 (1.3)	-0.5 (0.9)
Education (>12)	-	-	1.3 (3.6)	-	-0.2 (0.6)

Table 3 shows that there exists a significant non-interview bias: $bias_{NON}(p)=50.9\text{-}53.0=-2.1\%$. Expressed in relative terms we have $rbias_{NON}(p)=-4.1\%$ (the relative biases are in brackets). Similar conclusions can be drawn from Table 4: if we compare the population value ($P=51.5\%$) with the average for the initial respondents ($p_{RESP}=52.5\%$) of all SJM surveys, we have $rbias_{NON}(p)=-2.0\%$. Obviously, the first component of the non-response bias dominates in the case of gender. This is not surprising as we don't expect a considerable variation in the structure of gender across the clusters, so we have very small $bias^{(2)}_{NON}(p)$.

From Table 4 we can make a more precise estimate of the substitution bias. The difference between substitution and weighting imputation for non-interview cases is considerable: diff=54.6-52.3=2.3%. The relative bias in the whole sample is much smaller because of the relatively high proportion of respondents - we have $rbias^{(N)}_{SUB}(p)=0.7\%$. Thus, on average, the substitutions introduce a relative imprecision, close to one percent of the estimate. A possible interpretation would be that, for example, among substitute non-interview units a male non-interview occurs more often than among initial non-interview units. To evaluate the relative importance of this bias we consider the coefficient of variation in the combined sample of the 6 SJM surveys: we have $cv_{12,600}=0.9\%$. There, the sampling error shrinks by the factor $\sqrt{6}$ and the relative bias becomes comparably much larger. It is more correct, of course, to use the factor $F=1.6$ from section 2.3.2 to calculate the comparable - but still conservative - coefficients of variation: $cv'_{12,600}=0.6\%$ and $cv'_{2,100}=1.4\%$ respectively. We can conclude that the relative biases $rbias^{(N)}_{SUB}(p)$ and $rbias^{(2)}_{NON}(p)$ are not statistically significant. On the other hand, the non-interview bias $rbias_{NON}=-2.0\%$ is significant, at least in the context of 6 SJM surveys ($\alpha<0.05$). The same is true, of course, for the total substitution bias $rbias^{(T)}_{NON}(p)$.

Of course, we can not speak about classical hypothesis testing here, since all the assumptions needed are not fulfilled. There is a lack of a proper randomization when the missing units and substitute units are "selected". As a result, the substitution bias does not behave as a simple difference of two proportions. We can thus observe that this bias is highly consistent across the surveys, so it can not be neglected just because it is smaller than the sampling error. As mentioned, the estimates cv'(p) are conservative. This can be clearly demonstrated by the fact that we observe much less variation among the 6 SJM surveys when estimating relative biases than we would expect from the values of cv'(p) and cv(p). This holds for all the variables observed.

5.1.2 Age

As opposed to gender, there is only a weak non-interview bias in the age structure, and the figures in Table 2 are typical of all SJM surveys. Again, only the first component of the non-response bias is important; the respondents differ from non-respondents in the category of younger persons. The persons from the age group 18-30 years for example are more likely to be non-respondents: in Table 3 we have a $bias^{(1)}_{NON}(p)=0.3\%$. This is further underestimated by the use of substitutions and also by weighting as initial non-responding units with $p_{NON}=27.7\%$ were replaced with $p_{WGH\text{-}I}=p_{SUB\text{-}I}=25.8\%$. From Table 4 we have the estimate: $rbias^{(N)}_{SUB}(p)=1.5\%$. As with the gender, this component of the bias is smaller than the coefficient of variation ($cv_{12,600}=1.5\%$ and $cv'_{12,600}=0.9\%$).

5.1.3 Rurality (settlements smaller than 2000 inhabitants)

We can observe (Table 3) a considerable non-interview bias $rbias_{NON}(p)=-2.5\%$ towards persons from larger settlements. Different for age and gender, the second component of the non-response bias dominates here: the non-interviews occur more often in urban clusters, so $rbias_{NON}^{(2)}(p)=-3.2\%$ (Table 4). On the other hand, the responding and non-responding units within clusters do not differ - they are all from the same cluster - and consequently $rbias^{(1)}_{NON}(p)$ is small. Also, we can observe some consistent differences between substitutions and weighting. Among the non-interview units in Table 2 there are only $p_{NON}=47.7\%$ units from smaller settlements, but among substitutes the we have $p_{SUB\text{-}I}=49.5\%$. The implicit imputation of the weighting performs better - the proportion is $p_{SUB\text{-}W}=47.3\%$ and we have the $rbias^{(2)}_{NON}(p)=-2.1\%$ in SJM 91/2. Similarly, we can observe from Table 4, not only a non-interview $bias^{(2)}_{NON}=-3.2\%$, but also a very consistent substitution bias $rbias^{(N)}_{SUB}=-0.6\%$ which is not so small compared to the sampling variability: $cv_{12,600}=2.7\%$ and $cv'_{12,600}=0.5\%$.

Table 4: The relative substitution bias in control variables in SJM surveys

variable	initial respondents	implicit input. by weighting	substitution (respondents)	FINAL SAMPLE substitutions	FINAL SAMPLE weighting	$rbias^{(2)}_{NON}$ |5-1|/5*100	$rbias^{(N)}_{SUB}$ |5-4|/5*100
	1	2	3	4	5	6	7
Gender (F)							
SJM90/1	52.9	51.4	56.0	53.5	52.6	-0.6	-1.7
SJM90/2	52.4	50.4	53.8	52.7	52.0	-0.8	-1.4
SJM91/2	53.0	53.5	51.3	52.7	53.1	0.2	0.8
SJM92/1	52.9	53.8	54.3	53.2	53.1	0.4	-0.2
SJM92/2	50.3	52.0	53.4	50.8	50.6	0.6	-0.4
SJM93/1	53.5	52.9	58.8	54.5	53.4	-0.2	-2.1
$SJM_{12,600}$	52.5	52.3	54.6	52.9	52.5	0.0	-0.7
deff(p)=1, $se_{12,600}$=0.44%, $cv_{12,600}$=0.9%, $cv_{2,100}$=2.2%, F>1.6 --> $cv'=_{12,600}$=0.6%, $cv'_{2,100}$=1.4%							
Age 18-30							
SJM90/1	29.5	27.9	31.3	29.9	29.2	-1.0	-2.4
SJM90/2	24.8	21.3	27.9	25.4	24.1	-2.9	-5.4
SJM91/2	26.3	25.8	25.8	26.2	26.2	-0.4	-0.1
SJM92/1	24.0	23.5	25.4	24.3	23.9	-0.4	-1.7
SJM92/2	25.0	23.8	22.9	24.7	24.8	0.8	0.4
SJM93/1	27.1	27.7	28.2	27.3	27.2	-0.4	-0.4
$SJM_{12,600}$	26.1	25.0	26.9	26.3	25.9	-0.4	-1.5
deff(p)=1, $se_{12,600}$=0.38%, $cv_{12,600}$=1.5%, $cv_{2,100}$=3.7%, F>1.6 --> $cv'_{12,600}$=0.9%, $cv'_{2,100}$=2.3%							
Rural (<2000)							
SJM90/1	53.3	45.6	46.3	51.9	51.8	-2.9	-0.2
SJM90/2	48.5	42.4	41.8	47.2	47.1	-3.0	-0.2
SJM91/2	53.4	47.3	49.5	52.8	52.3	-2.1	-1.0
SJM92/1	49.7	38.3	38.3	47.2	47.2	-5.3	-0.1
SJM92/2	52.1	47.8	47.4	51.3	50.7	-2.8	-1.2
SJM93/1	52.9	43.5	44.7	51.4	51.2	-3.3	-0.4
$SJM_{12,600}$	51.7	44.2	44.7	50.3	50.1	-3.2	-0.6
deff(p)=10, $se_{12,600}$=1.4%, $cv_{12,600}$=2.7%, $cv_{2,100}$=6.7%, F>6 --> $cv'_{12,600}$=0.5%, $cv'_{2,100}$=0.9%							
Education (>12)							
SJM90/1	38.4	40.5	44.0	39.5	38.8	1.0	-1.8
SJM90/2	37.3	41.5	44.5	38.8	38.0	1.8	-2.1
SJM91/2	34.5	34.6	38.5	36.0	35.8	3.6	-0.6
SJM92/1	36.0	38.7	39.4	36.7	36.6	1.6	-0.3
SJM92/2	36.4	39.1	39.1	36.9	36.9	1.4	-0.1
SJM93/1	36.4	39.7	39.4	36.9	37.0	1.6	0.3
$SJM_{12,600}$	35.5	39.5	40.8	37.5	37.2	2.2	-0.8
deff(p)=2.5, $se_{12,600}$=0.7%, $cv_{12,600}$=1.9%, $cv_{2,100}$=4.6%, F>2.5 --> $cv'_{12,600}$=0.8%, $cv'_{2,100}$=1.8%							

5.1.4 Education

Regrettably, we do not have the population values for the education variable from the register; nevertheless, we know the population structure from the 1991 Census. There may exist a certain over-reporting of education in the SJM surveys, but we assume that there still remains a considerable non-interview effect. The percentage of persons with 12 years of education or more is p_{RESP}=34.5% among respondents (Table 2), which is higher than the P=32.3% from the Census. In Table 4 we also have $rbias^{(2)}_{NON}$=2.2%. What is more important is that, again, substitutions perform poorly in removing this bias - in Table 2 we see that the percentage is even worse - $p_{SUB-I}=38.5\%$. Of course, we assume[9] the percentage of persons with at least 12 years of education to be lower among the non-interview units (perhaps around 30%) than among responding units in order to compensate for the high percentage among the respondents. Here again, across all surveys in Table 4 we consistently observe not only a certain non-interview bias $rbias^{(2)}_{NON}=2.2\%$ - and at least a few percentage points of $rbias^{(1)}_{NON}$ - but also a consistent substitution bias, $rbias^{(N)}_{SUB}=0.8\%$ which is not negligible when compared to $cv'_{12,600}=0.8\%$.

Obviously, the non-interview bias arises from the within cluster difference between respondents and non-respondents. The substitutions and the (sample) weighting can not remove this bias, since the data are not missing completely at random within cluster. In fact, they even create an estimate that is less accurate than the estimate based on respondents.

5.2 Bias in target variables

With few exceptions (voting) the population parameters of the target variables are generally unknown. Thus, we can only estimate the substitution bias $rbias^{(N)}_{SUB}(p)$ and the second component of the non-interview bias $rbias^{(2)}_{NON}(p)$. As mentioned, to evaluate the substitution bias we assume the weighting procedure to be the proper adjustment. This can be additionally justified by the previous observations on the control variables where weighting performed uniformly better than substitutions for all control variables.

The relative substitution bias behaves similarly in all SJM surveys. Some of the typical results about attitudes can be observed in Table 5. As with the control variables, the substitution bias is relatively low, especially when compared to a sampling error based on a single SJM survey ($n \approx 2,100$) even if we adjust for the factor F. For the target variables this factor is around F=2, as the design effect is generally between deff=1.3 and deff=1.7.

Table 5: The (absolute) relative biases for some target variables (%)

variable	survey SJM	initial respondents	implicit input. by weighting	substitution (respondents)	FINAL substitutions	SAMPLE weight.	RELATIVE BIAS (2)NON	(N)SUB	cv(p)
happiness	91/2	47.2	39.9	48.8	47.5	45.8	3.0	3.7	2.9
nuclear plant	91/2	14.0	12.4	15.3	14.2	13.7	2.2	3.6	6.3
adultery	91/2	7.8	7.8	11.0	8.4	7.8	0.0	7.6	9.1
refugees	92/2	6.6	6.6	7.7	6.8	6.6	0.0	3.0	9.3
army	91/2	14.1	12.5	10.3	13.3	13.8	2.1	3.6	6.3
government	91/2	43.8	42.8	51.5	45.3	43.6	0.5	3.8	3.0
religion	91/2	15.1	14.5	17.3	15.5	15.0	0.6	3.3	5.8

[9] We assume this with great certainty as we are not completely clear about the reasons for the discrepancy between the Census and the SJM survey estimates.

The distribution of the relative biases for all 180 variables considered is shown in Table 6.

Table 6: Distribution of variables according to the $rbias^{(N)}_{SUB}$ in the SJM surveys

rbias(N)SUB	percentage of variables
0.0% - 0.5%	34%
0.5% - 1.0%	22%
1.0% - 2.0%	26%
2.0% - 5.0%	11%
5.0% - 9.0%	7%

It can be observed that the relative biases are small in comparison to the relative sampling errors (coefficient of variation). Also the ratio R=rbias(p)/cv(p)=bias(p)/se(p) is generally low. Among all 180 variables in Table 6, half of them have this ratio below R=0.5 and only 10 variables have a ratio above R=1, the maximum being at R=1.8. Of course, using cv'(p) the ratio is approximately two times larger. It should be stressed that the estimated non-interview bias $rbias^{(2)}_{NON}$ is consistently smaller than substitution bias $rbias^{(N)}_{SUB}$. However, the final effect of the bias depends on signs of the $rbias^{(N)}_{SUB}$: if it is of the same sign as $rbias^{(1)}_{NON}$ then the substitutions enlarge the bias compared to weighting and, in general, this will be the case. The reason for this is the fact that respondents from substitute units differ from non-respondents to a greater extent than initial respondents. Thus, in general, the weighting would give smaller bias than substitutions do. If also the term ($rbias^{(2)}_{NON}$ - $rbias^{(N)}_{SUB}$) is of the opposite sign, even[10] the estimate based on responding unit will have a smaller bias than the estimate based on substitutions.

5.3 The variance

When we compare the variance of the substitution procedure ($n \approx 2{,}100$) with the variance of the weighting adjustment ($r \approx 1{,}700$) we face the problem of a different sample sizes. We have to carefully remove this factor to obtain the comparable variances. First, the intra-cluster correlation has to be estimated from the sample with substitutions:

$$roh_S(p)=(deff_S(p) - 1)/(b_S - 1),$$

where b_S is the average cluster size. Next, the design effect is recalculated:

[10] We can observe this from the equation on the page 48.

$$deff'_s(p) = 1 + roh_s(p)(b_r - 1),$$

where b_r is the average cluster size in the sample of respondents. The ratio of two comparable variances var_{SUB}/var_{WGH} thus equals the ratio of two design effects $deff'_s(p)$ and $deff_w(p)$ which is the design effect from the weighted sample. Some results can be observed in Table 7, where the weighting was done at the PSU level.

Table 7: Variance calculations for some target variables in SJM 91/2

variable	FINAL SAMPLE substitutions	FINAL SAMPLE weighting	SUBSTITUTION BIAS (%)	SUBSTITUTION deff	SUBSTITUTION roh	SUBSTITUTION cv	WEIGHTING deff	WEIGHTING roh	WEIGHTING cv	ratio of variances var_{SUB}/var_{WGH}*100
happiness	47.5	45.8	3.7	1.6	4.0	2.9	1.4	3.4	3.1	105.7
nuclear plant	14.2	13.7	3.6	1.4	2.7	6.3	1.4	3.2	7.1	95.6
adultery	8.4	7.8	7.7	1.6	4.0	9.0	1.7	6.6	11.0	80.9
refugees	26.1	26.9	-3.0	1.6	4.4	4.7	1.4	3.8	4.8	105.4
army	13.3	13.8	-3.6	1.4	2.7	6.6	1.4	3.3	7.1	94.4
government	45.3	43.6	3.3	1.5	3.4	2.9	1.4	3.8	3.3	96.2
religion	15.5	15.0	3.3	1.3	2.0	5.8	1.1	1.2	6.2	109.9
Education	36.0	35.8	0.6	1.9	6.5	4.0	1.7	6.7	4.3	98.8
Rural	52.8	52.3	1.0	11.0	72.8	6.9	9.0	72.9	7.0	99.9
Age	26.2	26.2	0.0	1.0	0.2	3.7	1.0	0.0	4.1	102.2
Gender	52.7	53.1	-0.8	1.0	-0.1	2.1	1.1	1.0	2.4	86.9

There is no clear pattern in the ratio of the comparable variances. Obviously, this ratio varies considerably across the variables. However, we can calculate the median and the average (median) increase in the comparable variance for all 30 variables from each survey. If we summarize the extensive calculations for the 6 SJM surveys we end up with the following rough approximations:

- weighting at the PSU level ($b=15$) $\Rightarrow var_{SUB}(p) \approx var_{WGH}(p)$,
- weighting at the SSU level ($c=5$) $\Rightarrow var_{SUB}(p) \approx var_{WGH}(p)/1.1$.

The increase in the variance due to weighting is relatively small. In the case of weighting at the PSU level, where we would expect an increase of 2% (Table 1) the actual increase - as a general tendency of the average increase - is completely negligible. When the weighting adjustments are performed at the SSU level we would expect an increase of 12% (Table 1), but the actual average increase is around 10%, though it differs considerably across variables. Thus, some other properties of the weights are also involved here. In any case, we can conclude that this increase is not a serious drawback of the weighting procedure in SJM surveys. The substitutions thus lose their most important comparative advantage - the advantage of improved precision. Obviously, all the conditions needed (low response rate, high intra-cluster correlation, small clusters) for a comparably lower variance with substitutions are not fulfilled in the case of SJM surveys.

6 Discussion

a) There exists a significant non-interview bias and also a significant non-response bias in the SJM surveys. This is most evident in the case of gender, urban-rural settlement and education. It may be concluded with considerable certainty that the bias also exists in many target variables where there are no population parameters to control. This can be additionally supported by the fact that there exist correlations between the control and target variables. Thus the target variables should also be affected by the non-response bias. However, we found that the biases are relatively small, which is due to the following factors:

i) The non-response bias $Bias_{NON}(p)$ is the product of the proportion of non-respondents (W_{NON}) and the difference between the values of respondents and non-respondents (Cochran, 1977: 361):

$$Bias_{NON}(p) = W_{NON}(P_{RESP} - P_{NON}).$$

Similarly, the relative bias is: $Rbias_{NON}(p) = W_{NON}(P_{RESP} - P_{NON})/P$. Thus, for example, the response rate of 80% with a 10% absolute and 20% relative difference[11] between respondents and non-respondents would result in the $Bias_{NON}(p) = 2.0\%$ and the $Rias_{NON}(p) = 4.0\%$. The same is true for non-interview bias and also for substitution bias.

ii) Because of the small sample size, the biases are dominated by the sampling errors. Thus, within one SJM survey, the bias will hardly be larger than the sampling error.

b) The substitutions improve the estimates only in the case of the urban-rural component of the bias, but this improvement is smaller than with weighting. If we assume that - due to the large intra-cluster correlation - the characteristics are closely related within certain areas, then these adjustments would, in general, be beneficial. On the other hand, however, there is clear evidence that these adjustments are not very helpful for other control variables. On the contrary, these adjustments often deteriorate the estimates. There is a clear evidence of substitution bias with all the control variables, especially with rurality and education; furthermore, with many important target variables, this relative (net) substitution bias can reach up to 7% of the estimates, and the total of the substitution bias $rbias^{(T)}_{SUB}$ can be even larger. Of course, similar to the non-interview bias, the substitution bias is relatively low in comparison with the sampling errors. It is, therefore, up to the users to decide how acceptable it is to

[11] Which is a large difference to be found in practice; for example we have this in the case of gender: $p_{NON}=0.42$ and $p_{RESP}=0.52$.

have a consistent substitution bias of a few percentage points, as in some control and target variables. Of course, the statistical significance does not necessarily indicate the substantial importance of the difference. On the other hand, however, a serious distortion in the bias is quite possible without statistical significance.

c) Weighting at the PSU level gave a relative mean squared error:

$$RMSE_{WGH}(p) = CV^2(p)(1+VIF(p)) \approx CV^2(p),$$

since $VIF(p) \approx 0$. At the SSU level we have $VIF(p) \approx 0.1$, an increase of 10% in $RMSE_{WGH}(p)$. There[12] is no extra bias since the bias is the same as in the case of PSU weighting.

In case of substitutions we have:

$$RMSE_{SUB}(p) = CV^2(p)(1 + (Rbias^N{}_{SUB}(p))^2/CV^2(p)).$$

The estimate for the median ratio for target variables was $R=rbias(p)/cv(p)=0.5$ which gives the increase of $R^2=rbias^2(p)/cv^2(p)=25\%$. Of course, for some variables this increase may even be above 100%. Obviously, the weighting at the PSU level gives the lowest increase in the MSE(p). The other two options - substitution procedure and the weighting at the SSU level - are much worse. Of course, the substitutions additionally involve extra field-work operations.

In case where there is no adjustment, but a comparable sample size of initial respondents (for this we would need a sample of $n \approx 2636$), using a ratio estimator we would implicitly assume that the sample is self-weighted, which is not the case. Such an estimator is thus not correct. However, in practice it is used very often and experience shows that variance remains approximately the same as in cases with no non-response (or in a sample with substitutions with comparable sample size). Thus $VIF(p) \approx 0$. Instead, however, we have an increase in the bias, but $bias^{(2)}{}_{NON}(p)$ is generally (in absolute terms) smaller that $bias^{(N)}{}_{SUB}(p)$. Only in the case of the variables connected with rurality is the $bias^{(N)}{}_{SUB}(p)$ smaller than $Rbias^{(2)}{}_{NON}(p)/CV(p)$. We have the median ratio $Rbias^{(2)}{}_{NON}(p)/CV(p)= 0.25$. Of course, the comparison is more complex because of the issue of the sign of the biases. Nevertheless, in general with a sample of respondents the median increase in $RMSE_{RESP}(p)$ would be about 6% $(0.25^2=0.63)$ what is smaller than in the case of the weighting at the PSU level. Generally speaking, the sample of respondents gives the lowest increase in RMSE(p) for the set of 180 variables

[12] This is an empirical fact.

when it is assumed that the weighting procedure has no bias. However, for many variables (linked with the rurality) this is definitely not the case.

d) The bias-variance considerations contribute nothing radically new to the evaluation of the substitutions in the SJM surveys. This is mainly due to the existing setting of the SJM survey:

- a small sample size, keeping the biases small in comparison to the sampling errors;
- a focus on the "soft" attitude variables, which are less sensitive to non-interviews compared to more factual variables, e.g. unemployment, income;
- a good sampling frame which not only enables a proper probability sample but also maintains field-work controls over substitutions;
- a relatively low non-response rate and, as a consequence, relatively low biases;
- a relatively large PSUs ($b=15$) keeping the increase in the variance due to the weighting to an almost negligible level.

e) Clearly, the units within the clusters are not always missing completely at random (MCAR), so substitutions and weighting adjustments - both assuming MCAR within clusters - can remove only the second component of the non-interview bias.

It is very likely that both approaches to non-response are inadequate and some population weighting (raking, postratification, linear weighting) or even some explicit modelling (Brehm, 1993) of the non-response would give better results. But again, since the non-response bias is generally small (at least when compared to the sampling error) in SJM surveys, the gains from this procedures might be also very small. In any case, some sort of population weighting using the available auxiliary variables (gender, age, rurality) should be explored to establish the optimal strategy for the non-response problem in the SJM survey.

f) There does, of course, exist the problem of non-eligible units. Throughout our analysis we have been dealing with the non-interviews and not with the non-responses as we have substituted and weighted for the non-eligible units and not for the non-responding units. We can additionally justify this approach by the following:

i) In all important aspects, the non-eligible units behave similarly to non-responding ones: in later waves of substitutions the proportions of both categories increase and both categories behave in the same way with regard to the control variables.

ii) A separate calculation was performed for the SJM 91/2 survey where the non-eligible units - and also the corresponding substitutions - were excluded from the calculations. The main finding was that all the above conclusions firmly hold, although to a slightly lesser extent, which was due to the smaller proportion of missing cases and to the smaller sample.

iii) The non-eligible units in the SJM survey are often non-eligible only within a specific cluster and the majority of the SJM non-eligible units are in fact eligible citizens (voters) of Slovenia since the register of population effectively serves many important purposes (taxes, voting lists, certificates). Consequently, it might be more appropriate to treat them as non-respondents - in this case the above compensation procedures (substitution, weighting) can be justified.

7 Conclusions

a) Field substitutions are a specific and controversial solution to the unit non-response problem. The proper evaluation of this procedure should consider not only the field-work process and bias-variance issues, but also the general milieu of the survey: the nature of the sampling frame, costs, survey tradition, timing of the survey and the needs of the users.

b) The impression is that the apparent simplicity of substitutions being sufficient reason for their use is becoming increasingly questionable. This is especially true if we are fully aware of some serious drawbacks: problems with the field-work controls, a considerable prolongation of the field-work and a possible decrease in the response rate, at least in the face-to-face surveys. Thus, the remaining advantage of the substitution procedure is that of increased precision, because the designed structure of the sample is obtained. However, for this advantage to become a reality some specific circumstances have to exist. Evidently, they do not exist in the SJM survey.

c) Some conclusions concerning the SJM survey:

i) All the serious disadvantages of the substitution procedure hold very firmly for the SJM survey.

ii) The advantage of apparent simplicity to the users (self-weighted sample) also holds to some extent. Similarly, the continuity of comparable methodology is also a very significant factor.

iii) There is only a negligible decrease in precision when alternative weighting is used, which is mainly due to the specific setting of the SJM surveys. Nevertheless, because of the bias we still have a slightly lower mean squared error with weighting procedure compared to the sample with substitutions. It is somehow surprising, however, that the (non-weighted) sample of respondents has, in general, the smallest mean squared error.

iv) There exists a special bias - up to few percentage points of the estimates - additional to the non-response bias, which is introduced by the substitution procedure. This bias is generally higher than the bias arising from the unweighted sample of respondents. However, all those biases are generally dominated by the sampling error.

d) As the bias-variance considerations (*iii*) contribute relatively little to the evaluation of the substitutions, the answer to the initial question - whether the substitutions in the SJM survey can be justified - lies somewhere between the first two (*i* and *ii*) arguments above. However, the argument of simplicity relies solely on the fact that by substitutions we avoid weights and/or other non-response adjustments. But we have already observed that substitutions have a relatively small effect on non-response bias. Often, the substitutions additionally deteriorate the estimates. So there still exists a need for proper non-response adjustments. Of course, when the non-response bias is relatively small - as in SJM surveys - we can simply ignore this bias. But, to ignore a non-response problem, why should we bother with substitutions? For, when no non-response adjustments are made, the substitutions are simply redundant[13]. Interestingly enough, even the author of the SJM methodology thirty years ago felt disquiet of the effects of the substitution procedure as he observed some unexplainable distortions in the control variables (Blejec, 1965:43).

References

[1] Blejec M. (1970): *Slovensko javno mnenje 1969*, Ljubljana: VŠSPN.

[2] Blejec M. (1965): *Plan in analiza vzorca za anketo o masovnih komunikacijskih sredstvih 1962*, Ljubljana: VŠSPN.

[3] 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.

[4] Brehm J. (1993): *Phantom Respondents. Ann Arbor: University of Michigan Press.*

[5] Chapman D. (1983): The Impact of Substitutions on Survey Estimates. In W.G. Meadow, I. Olkin, and D. Rubin (Eds.): *Incomplete Data in Sample*

[13] The General Social Survey in the US practiced this simple solution: the initial sample was increased according to the expected non-interview rate. After this no substitutions were used and no other adjustments were employed.

Surveys, **Vol.II**, Theory and Bibliographies. New York: Academic Press, 45-61.

[6] Chapman D. (1986): Substitution for Non-response in Sample Surveys. Paper resented at a Statistical Methodology Meeting, NCHS and Washington Statistical Society.

[7] Chapman D., and Roman A.M (1985): An Investigation of Substitution for an RDD Survey. Proceeding of the Survey Research Methodology Section, American Statistical Association, 269-274.

[8] Cochran W. (1977): *Sampling Techniques*, New York: Wiley.

[9] Falorsi, Russo (1993): *Unit non-response in a two stage cluster sample.* Contributed paper. Firenze: ISI.

[10] Foreman E. (1991): *Survey Sampling Principles*, Marcel Dekker Inc.

[11] Forsman G., and Berg S. (1992): *Telephone Interviewing and Data Quality, An Overview and Empirical Study*. Linkoping University..

[12] Fuller W. (1989): *PC_Carp*. Statistical Laboratory, Iowa State University.

[13] Groves R. (1989): *Survey Errors and Survey Costs*, New York: Wiley.

[14] Hansen M., Hurwitz W., and Madow W. (1953): Sample Survey Methods and Theory, Vol I, Methods and Applications. New York: Wiley.

[15] Kalton G., and Moser (1971): *Survey Methods in Social Investigation.* London: Heineman Education Book.

[16] Kalton G.: (1983): *Compensating for Missing Survey Data.* Ann Arbor: ISR.

[17] Kish L. (1965): *Survey Sampling*. New York: Wiley.

[18] Kish L. (1992): Weighting for unequal Pi. *Journal of Official Statistics*, **8**, 183-201.

[19] Judkins D. (1991): Non-response adjustment for a telephone follow-up to a national in-person survey. Proceedings of the 1991 ASA Survey Research Methods Section, 584-586.

[20] Lessler J. (1991): *Nonsampling Errors in Survey*. New York: Wiley.

[21] Little R., and Rubin D. (1987): *Statistical Analysis with Missing Data.* New York: Wiley.

[22] Little R., and Rubin D.B. (1989): The Analysis of Social Science Data with Missing Values. *Sociological Methods and Research*, **18**, 292-327.

[23] Madow G., and Olkin I. (1983): *Incomplete Data in Sample Surveys*. Vol. III/Proceedings of Symposium. New York: Academic Press, 494.

[24] Marliani G., and Pacei S. (1993): Effects of Household Substitution in the Italian Consumer Expenditure Survey. *Bulletin of the International Statistical Institute*, Contributed Paper, 49th ISI Session, Firenze:ISI, 149-150.

[25] Nathan G. (1980): *Substitution for Non-response as a Means to Control Sample Size. Sankhya*, **C42**, 50-55.

[26] *Report on 4th International Workshop on Household Survey Non-Response* (1993), Bath, 7-9 September 1993.

[27] Sardnal C., Swensson B., and Wretman J. (1991): *Model Assisted Survey Sampling*. New York: Springer-Verlag.

[28] United Nations (1982): National Household Survey Capability Programme, Non-Sampling Errors in Household Surveys: *Sources, Assessment and Control*. New York.

[29] Vehovar V. (1993): The Field Substitution in the Slovene Labour Force Survey. *Bulletin of the International Statistical Institute*, Contributed Paper, 49th ISI Session, Firenze: ISI, 519-520.

[30] Verma V. (1992): *Household Surveys in Europe: Some Issues in Comparative Methodology*. Seminar: International Comparison of Survey Methodologies, Athens 30 March-1st April 1992, pp. 17.

[31] Verma V., and Gabilondo L.G. (1993): *Family Budget Surveys in the EC: Methodology and Recommendations for Harmonization*, Eurostat, Statistical Document 3 E, Luxembourg, pp. 136.

[32] Yates F. (1960): *Sampling methods for censuses and surveys*. London: Charles Griffin.