TUTKIMUKSIA FORSKNINGSRAPPORTER RESEARCH REPORTS 219

INTERNATIONAL PERSPECTIVES ON NONRESPONSE

PROCEEDINGS OF THE SIXTH INTERNATIONAL WORKSHOP ON HOUSEHOLD SURVEY NONRESPONSE

SEPPO LAAKSONEN (ED.)

Tilastokeskus
Statistikcentralen
Statistics Finland

TUTKIMUKSIA FORSKNINGSRAPPORTER RESEARCH REPORTS 219

International Perspectives on Nonresponse

Proceedings of the Sixth International Workshop on Household Survey Nonresponse 25–27 October 1995

Edited by Seppo Laaksonen

Statistics Finland, Helsinki

Cover design
Maija Sohlman

Lay-out
Tiina Isaksson

ISSN 0355-2071
ISBN 951-727-217-0

OY EDITA AB
Helsinki 1996

PREFACE: NONRESPONSE – AN ESSENTIAL INDICATOR OF SURVEY QUALITY

Statistics Finland was charged with the challenging task of arranging the Sixth International Workshop on Household Survey Nonresponse last year. The workshop was held at the premises of Statistics Finland in late October 1995. Ever since the fairly free-format first meeting in Stockholm in September 1990 an intimate group of enthusiastic multi-scientific researchers and survey statisticians has come together in Washington, Voorburg, Bath and Ottawa; thus these meetings can soon be regarded as a tradition. I would like to take this opportunity to thank all those who have actively contributed to the success of the workshops. Right from the very first meeting in Stockholm Bob Groves and Lars Lyberg in particular have been among the most enthusiastic to promote this tradition.

While a number of excellent papers, typically preliminary, were presented, the workshops provided a good opportunity for innovative and deep discussions in a relaxed atmosphere thanks to their informal nature. The participants in the teams are volunteers who are seriously worried about the high survey nonresponse in their countries and try to do their best in developing the quality of surveys. It has become more and more clear that nonresponse should be considered as an essential indicator of survey quality. In some countries, the information and nature of nonresponse are an actual part of the quality system of survey institutes. I presume that this will be the future trend.

Due to the earlier mentioned informal nature of the workshops no printed workshop publications have been produced so far, but short workshop reports instead. This proceedings is an exception. We will see whether this will be a one-off occasion or becoming a tradition as well. We proposed to the participants of the workshop that a shortened and more finalised version of the workshop papers be published and as we met with reasonable response, we decided to move ahead with the publication. It comprises the versions of most workshop papers. In addition, the publication also includes the first Finnish 'nonresponse' barometer as an appendix. Many parts of this appendix were presented in a poster session.

The proceedings starts with the invited paper of Carl-Erik Särndal. He was a visiting researcher and consultant at Statistics Finland at the time of the workshop. We are very happy to be able to publish his interesting paper in this proceedings. We believe that the paper will be often referred to, since it develops a new synthesis between reweighting methods and imputation methods.

The major part, i.e. the middle part of the proceedings consists of other available papers. These are mostly papers focusing on pre-survey preparations and survey designs, whereas the rest of the papers highlight

post-survey adjustments. All the papers can be considered 'international perspectives on nonresponse,' and thus the heading title of the proceedings. I would like to thank Mick Couper for this suggestion.

Developing the text for the heading was a real problem and therefore, I made a survey where some titles were proposed and I asked the opinion of sampled persons on these suggestions. In addition, new suggestions were invited. I drew a 30 % intentional sample of the workshop participants and sent the questionnaire to them (note that no pre-tests were used). The response rate was about 70 %. All the answers were acceptable, although there was a varying degree of 'carefulness'.

The second best proposal, basically a reformulation of an original proposal, was chosen as the title of this preface. Many other suggestions also illustrated the nature of the workshop and the papers extremely well, and therefore, it is appropriate to list them here:

'Dealing with nonresponse from pre-survey preparations to post-survey adjustments'
'Multi-metholodogical approaches to nonresponse research'
'Need for Total Nonresponse Management'
'New Directions in Managing Nonresponse'
'Non-response, research and measures.'
'Handling nonresponse: theory and practice'
'Nonresponse management, aims and means.'

If you are not satisfied with my solution you are free to make your own choice.

***

The workshop itself and the edition of the proceedings required a lot of work of several persons. I would like to thank all the participants, including the speakers, chairpersons and rapporteurs of the sessions, for their contributions. In particular, I wish to express my gratitude to Pertti Kangassalo, Lilli Japec, Kari Djerf, Markku Heiskanen, Markku Lindqvist and Paavo Väisänen on their excellent output in the organising of the workshop. Moreover, the edition of the proceedings would not have succeeded without the help of Tiina Isaksson and Marjo Koponen. Thank you very much for your efforts.

Helsinki, July 1996

Seppo Laaksonen
Editor

CONTENTS

FOR A BETTER UNDERSTANDING OF IMPUTATION

Carl-Erik Särndal
Université de Montréal

This paper proposes a common framework for analyzing the estimation error following from the two standard techniques for nonresponse treatment: seweighting and imputation. The proposed common framework recognizes the desire, present in both techniques, to restore (on the average) the estimator that one would have used with 100 % response; the estimator finally used thus becomes a "surrogate estimator". We discuss variance components (sampling variance, surrogate variance) as well as bias (surrogate bias). Several examples are given to illustrate the theoretical concepts.

Key words: Nonresponse bias, reweighting, imputation, variance components.

1. Introduction

Missing data occur in almost all surveys. Standard texts on survey theory mention two principal avenues for handling this problem at the estimation stage: reweighting and imputation. Reweighting consists of raising the original weights for the respondent (nonmissing) values when estimates are computed. Imputation implies creating plausible (but artificial) substitute values for all those missing, while preserving the original weights when estimates are calculated.

Imputation results in a complete rectangular data matrix. This significant practical advantage is perhaps the strongest argument in favour of imputation. Uniform treatment of all variables of interest becomes possible, using a predefined weight system.

It is sometimes suggested that imputation should be reserved for the item nonresponse and that the unit nonresponse be treated through reweighting. Clearly one may consider imputation for the unit nonresponse, too. The objective of a complete rectangular data matrix is realized whether imputation is limited to the item nonresponse, or whether it is extended to also cover unit nonresponse. The more extensive usage leads to a greater reliance on artificial values. Some may argue that this could do further harm to the estimates whose quality is already in question because of item

nonresponse. This is not necessarily so, but there is no clear consensus as to whether the more limited or the extended reliance on imputation should prevail. The question merits careful consideration.

Edit and imputation precede estimation in the statistical production process. One task of editing is to check the consistency of reponses given to a set of related items. The edit procedure checks each record using a predefined set of edit rules. When an inconsistency is encountered, the record fails the edit, and one or more of the recorded responses are set to missing. Consequently, "missing values" include those missing because of absence of response as well as those set to missing by the edit system. In the following, we use "nonresponse" as an all inclusive term for "missing". For example, "nonresponse units" include those with an absence of response as well as those with values set to missing by the edit procedure.

That nonresponse and missing data have harmful effects on survey quality is well known. In many countries, nonresponse rates are on the rise and are unlikely to decrease substantially in the future. To adopt fixed goal standards such as "nonresponse must not exceed 25 % in any of our surveys" seem unrealistic. Nonresponse varies with the nature of the survey. Also, it depends on the survey whether a rate of, say, 26 % missing is really so serious that the resulting quality is "catastrophically poor". In any case, survey statisticians always seem to adjust to new, higher levels of nonresponse, despite concerns voiced during a period of significantly rising levels. Rarely if ever does one hear of a decision to scrap the results of a survey because of unacceptably high nonresponse, although such a decision would no doubt be warranted in many cases.

Making the best of the situation will no doubt continue to be the strategy. In this perspective it is important to gain a better understanding of the procedures in use, in particular of imputation. This paper promotes an integrated view of reweighting and imputation. We use classical survey sampling concepts to gain a common perspective on the two means of treating nonresponse. Section 4 looks at reweighting; Sections 5 to 9 discuss imputation methodology.

2. Notation

Let $U = \{1, ..., k, ..., N\}$ be the index set for a set of N units. We can think of U as representing either the whole target population or a stratum of this population if the survey uses a stratified sampling design. Denote by y_k the true value for the unit k of a variable of interest y. We want to estimate the total $Y = \Sigma_U y_k$.

To assess the properties (bias, variance) of an estimator of Y, we keep the probabilistic basis as simple and as close to traditional survey sampling

theory as possible. The only two sources of randomness are the (known) sampling design used to select a sample, and the (unknown) response mechanism, that is, the probability distribution according to which nonresponse (missing values) occurs. An assumption is necessary about the form of the latter, since it is unknown. This set-up amounts to a traditional two-phase sampling scheme, except for the fact that in traditional two-phase sampling, the statistician knows and controls the second phase (as well as the first phase) selection.

The probability sample s is drawn from U with the *sampling design p(s)*. For this design, let π_k and $\pi_{k\ell}$ denote the known inclusion probabilities of first and second order. Then nonresponse occurs (by failure to respond or by failed edit). The respondent set is denoted r, the set of nonrespondents $o = s - r$. The y-data $\{y_k: k \in r\}$ are observed; the y-data $\{y_k: k \in o\}$ are missing. The *response mechanism*, denoted $q(r \mid s)$, expresses the unknown conditional probability that the set r is realized, given s. The corresponding unknown probability of response of k is denoted θ_k. The simplest response mechanism is the uniform mechanism which assumes that units respond independently of each other and with the same (but unknown) response probability, $\theta_k = \theta$ for all k.

The survey design, conceived before any sampling is carried out, prescribes an estimator of Y that would be used for full response. That is, the statistician has fixed a weight system, $\{w_k\}$, which is efficient and suitable under the conditions and which would yield $\hat{Y}_s = \Sigma_s w_k y_k$ as the full response estimator of Y. For example, $w_k = 1/\pi_k$ defines the weight system if the design calls for the Horvitz–Thompson (HT) estimator. We assume that the full response estimator $\hat{Y}_s$ is design unbiased, so that

$$E_p(\hat{Y}_s) - Y = 0 \qquad (2.1)$$

holds exactly as for the HT estimator or to a very close approximation as when $\hat{Y}_s$ is the ratio estimator or some other member of the generalized regression estimator (GREG) family. (For simplicity the GREG estimator is treated as design unbiased in the following.) Here E_p denotes expectation with respect to the sampling design *p(s)*.

Now nonresponse occurs, and a method for handling it becomes necessary. Reweighting and/or imputation enter the picture. Although often presented as distinct methods, reweighting and imputation have aspects in common. In both cases, a surrogate estimator, denoted here by $\hat{Y}_{*s}$, replaces the full response estimator $\hat{Y}_s$ which cannot be computed because of missing data. The statistician hopes or believes that surrogate and full response estimators do not differ systematically. The statistician considers the conditional expected value of $\hat{Y}_{*s} - \hat{Y}_s$, given s, to be zero (although he/she may not invoke these exact mathematical properties). Otherwise the

procedure would be hard to defend; certainly the statistician would not willingly introduce systematic error in the estimation.

Whether reweighting or imputation is used, an analysis of the variability starts from decomposing of the total error of $\hat{Y}_{*s}$ into sampling error and surrogate procedure error:

$$\hat{Y}_{*s} - Y = (\hat{Y}_s - Y) + (\hat{Y}_{*s} - \hat{Y}_s) \tag{2.2}$$

where $\hat{Y}_s$ is the full response estimator and $\hat{Y}_{*s}$ the corresponding surrogate estimator. Then

$$\mathrm{Bias}_{pq}(\hat{Y}_{*s}) = \mathrm{E}_{pq}(\hat{Y}_{*s}) - Y = \mathrm{E}_p\{\mathrm{B}_{\mathrm{SUR}|s}\} \tag{2.3}$$

where

$$\mathrm{B}_{\mathrm{SUR}|s} = \mathrm{E}_q(\hat{Y}_{*s} - \hat{Y}_s \mid s)$$

is the conditional expectation of the surrogate procedure error $\hat{Y}_{*s} - \hat{Y}_s$. As we have argued, the statistician likes to believe that

$$\mathrm{B}_{\mathrm{SUR}|s} = \mathrm{E}_q(\hat{Y}_{*s} - \hat{Y}_s \mid s) \approx 0 \tag{2.4}$$

If (2.4) holds, the surrogate procedure (reweighting or imputation) creates no bias, which is the main concern. It is true that nonresponse usually causes some variance increase as well, but it can be relatively easily tolerated, because it is measurable from the observed data whereas the bias is not.

The idea that, whatever sample s was drawn, the estimation is "patched up" by the surrogate procedure finds its mathematical expression through (2.4). The francophone term "redressement", frequently used in connection with nonresponse adjustment, expresses well the idea of "patching up" the situation.

To decompose the total error of $\hat{Y}_{*s}$ in the manner of (2.1) corresponds to the analysis of total error in traditional two-phase sampling. In such a design, the first phase consists of selection of a sample s for which one would like to observe the data $\{y_k: k \in s\}$. But since this may turn out to be expensive, some less expensive data are instead gathered. The second phase consists of a subselection from s of a considerably smaller subset r for which one records the y-data $\{y_k: k \in r\}$.

3. Error Decomposition

The decomposition (2.1) forms the starting point for analyzing the variance and other essential properties of the surrogate estimator $\hat{Y}_{\bullet s}$. We obtain

$$\mathrm{MSE}_{pq}(\hat{Y}_{\bullet s}) = \mathrm{V}_{pq}(\hat{Y}_{\bullet s}) + \{\mathrm{Bias}_{pq}(\hat{Y}_{\bullet s})\}^2 \qquad (3.1)$$

where $\mathrm{V}_{pq}(\hat{Y}_{\bullet s})$ and $\mathrm{Bias}_{pq}(\hat{Y}_{\bullet s})$ denote the variance and the bias of $\hat{Y}_{\bullet s}$, jointly with respect to the sampling design p and the response mechanism q.

To evaluate $\mathrm{V}_{pq}(\hat{Y}_{\bullet s})$, condition on s and use the customary decomposition $\mathrm{V}_p\mathrm{E}_q(\hat{Y}_{\bullet s} \mid s) + \mathrm{V}_p\mathrm{E}_q(\hat{Y}_{\bullet s} \mid s)$. Given s, define the conditional expectation and variance of the surrogate error $\hat{Y}_{\bullet s} - \hat{Y}_s$ as $\mathrm{B}_{\mathrm{SUR}|s} = \mathrm{E}_q(\hat{Y}_{\bullet s} - \hat{Y}_s \mid s)$ and $\mathrm{V}_{\mathrm{SUR}|s} = \mathrm{V}_q(\hat{Y}_{\bullet s} - \hat{Y}_s \mid s)$, respectively. Then the variance term in (3.1) can be expressed as

$$\mathrm{V}_{pq}(\hat{Y}_{\bullet s}) = \mathrm{V}_p(\hat{Y}_s + \mathrm{B}_{\mathrm{SUR}|s}) + \mathrm{E}_p(\mathrm{V}_{\mathrm{SUR}|s})$$

and the squared bias term can be expressed by (2.3). A simple development now gives

$$\mathrm{MSE}_{pq}(\hat{Y}_{\bullet s}) = \mathrm{V}_{\mathrm{SAM}} + \mathrm{V}_{\mathrm{SUR}} + 2\,\mathrm{Cov}_p(\hat{Y}_s, \mathrm{B}_{\mathrm{SUR}|s}) + \mathrm{E}_p(\mathrm{B}^2{}_{\mathrm{SUR}|s}) \qquad (3.2)$$

where $\mathrm{V}_{\mathrm{SAM}} = \mathrm{V}_p(\hat{Y}_s)$ and $\mathrm{V}_{\mathrm{SUR}} = \mathrm{E}_p\{\mathrm{V}_{\mathrm{SUR}|s}\}$. It is a variance component decomposition somewhat similar to the one resulting from the wellknown U.S. Bureau of the Census survey error model. The first two terms on the right hand side of (3.2) are easy to interpret: $\mathrm{V}_{\mathrm{SAM}}$ represents sampling variance, and $\mathrm{V}_{\mathrm{SUR}}$ represents variance added by the surrogate procedure. The third and fourth terms involve the conditional bias $\mathrm{B}_{\mathrm{SUR}|s}$. The covariance term may be negative; the other three are nonnegative.

In general, $\mathrm{B}_{\mathrm{SUR}|s} \neq 0$, and the MSE will contain four non-zero terms. However, in the special case when $\mathrm{B}_{\mathrm{SUR}|s} = 0$, (3.2) simplifies into

$$\mathrm{V}_{pq}(\hat{Y}_{\bullet s}) = \mathrm{V}_{\mathrm{SAM}} + \mathrm{V}_{\mathrm{SUR}} \qquad (3.3)$$

Using the MSE decomposition (3.2) and its special case (3.3), we discuss first reweighting in Section 4, then imputation in Sections 5 to 10.

4. Reweighting

Consider first reweighting. Starting from an original weight system $\{w_k\}$, where the w_k may depend on s and on auxiliary information (but not on the response set r), a surrogate estimator of Y can be constructed as

$$\hat{Y}_{\bullet s} = \Sigma_r w_k y_k / \hat{\theta}_k \qquad (4.1)$$

where $\hat{\theta}_k$ is an estimate of the unknown response probability θ_k. The logic of this procedure is obvious: we change the weight of unit k from its original value w_k to the increased value $w_k/\hat{\theta}_k$, thus compensating for the values missing. In particular, if the initial weight system is defined by $w_k = 1/\pi_k$, then (4.1) is the *reweighted Horvitz–Thompson estimator.*

A simple (but not very refined) reweighting is based on the inverse response rate, so that $1/\hat{\theta}_k = n/m$ for all k, where n and m are the respective sizes of s and r. Under a uniform response mechanism, the approach is unbiased: it is easy to see that $B_{SUR|s} = E_q(\hat{Y}_{\bullet s} - \hat{Y}_s \mid s) = 0$ holds. In practice, reweighting is sometimes routinely carried out using $1/\hat{\theta}_k = n/m$. The procedure has one practical advantage which may account for its popularity: in estimating the variance one can simply proceed (assuming that s is realized by simple random sampling and that the response mechanism uniform) as if the m respondents are a simple random sample from N.

Several papers propose more sophisticated approaches for estimating the θ_k. These make a more extensive use of auxiliary information. An example is Ekholm and Laaksonen (1991), who build an estimator for the Finnish Household Budget Survey. They start essentially from the Horvitz–Thompson weights $w_k = 1/\pi_k$, then model the response probabilities by logistic regression on a vector of household characteristics x_k observable for the sample households $k \in s$, that is, it is assumed that $\log\{\theta_k/(1 - \theta_k)\} = x_k'\beta$ well describes the reponse pattern.

Reweighting can take various other forms than the one implicit in (4.1). For example, instead of (4.1), consider another natural surrogate estimator, $\hat{Y}_{\bullet s} = (N/\hat{N})\Sigma_r w_k y_k/\hat{\theta}_k$, where $\hat{N} = \Sigma_r w_k/\hat{\theta}_k$. Here the original weight of y_k, which is w_k, is being transformed into $w_k(N/\hat{N})/\hat{\theta}_k$.

When auxiliary information is available, the estimator planned for full response case may be the GREG estimator. The starting weights w_k in (4.1) then depend on the auxiliary information available. They are given by $w_k = a_k g_k$, where $a_k = 1/\pi_k$ is the sampling weight and g_k is the weight adjustment (g-weight) calculated with the aid of the auxiliary information (known auxiliary totals) as described in Estevao, Hidiroglou and Särndal (1995).

Bethlehem (1988) shows that regression estimators are less affected by nonresponse bias than the Horvitz–Thompson estimator. Reweighting for regression estimation is also examined in Särndal and Swensson (1987), Särndal, Swensson and Wretman (1992), chapters 9 and 15, where estimates of the two components of variance are developed using the two-phase selection argument, assuming that the response mechanism (the second phase selection) follows the response homogeneity group (RHG) model. This model states that, in each of H sample subgroups, units respond

independently and with the same unknown response probability. The resulting reweighting factors become $1/\hat{\theta}_k = n_h/m_h$ for all k in subgroup h, $h = 1, ..., H$, where m_h/n_h is the response rate in group h. Simple explicit estimators are obtained for the two variance components, as illustrated in Example 4.1 which follows. Another example of reweighting is given in Binder (1991), who discusses weighting in analyzing categorical survey data with nonresponse.

Example 4.1. Estimation of variance components for reweighting by groups. Consider simple random sampling without replacement (SRSWOR) with the sampling fraction $f = n/N$. Then $w_k = 1/f$ for all k. The full response estimator is $\hat{Y}_s = \Sigma_s w_k y_k = N\bar{y}_s$ with the wellknown variance $V_{SAM} = N^2(1/n - 1/N)S^2_{yU}$ where $S^2_{yU} = \Sigma_U(y_k - \bar{y}_U)^2/(N - 1)$. (Here and in the following, overbar indicates arithmetic mean of the variable in question over the set of units appearing as an index, that is, $\bar{y}_U = \Sigma_U y_k/N$, $\bar{y}_s = \Sigma_s y_k/n$, $\bar{y}_r = \Sigma_r y_k/m$, $\bar{x}_r = \Sigma_r x_k/m$, and so on.) To help the nonresponse treatment, suppose that for every $k \in s$, we can observe an auxiliary variable value x_k as well as membership in one of H possible groups (the RHG groups, indexed $h = 1, ..., H$). Then a surrogate estimator is given by $\hat{Y}_{*s} = (N/n)(\Sigma_s x_k)B_r$ with $B_r = (\Sigma_{h=1}^{H} n_h\bar{y}_{r_h})/(\Sigma_{h=1}^{H} n_h\bar{x}_{r_h})$. Here, $\hat{Y}_{*s}$ is essentially without bias if the assumed RHG model holds, because, as is easy to verify, the conditional bias satisfies $E_q(\hat{Y}_{*s} - \hat{Y}_s \mid s) \approx 0$. The variance now consists of V_{SAM} plus a reweighting component, V_{REW}, so the total variance is estimated as $\hat{V} = \hat{V}_{SAM} + \hat{V}_{REW}$, where $\hat{V}_{REW}$ has the "conditionally post-stratified" form

$$\hat{V}_{REW} = N^2\Sigma_{h=1}^{H}(n_h/n)^2\,(1/m_h - 1/n_h)S^2_{er_h} \qquad (4.2)$$

where $S^2_{er_h} = \Sigma_{r_h}(e_k - \bar{e}_{r_h})^2/(m_h - 1)$ with $\bar{e}_{r_h} = \Sigma_{r_h} e_k/m_h$ and $e_k = y_k - \hat{B}_r x_k$. We omit the (more complex) expression for V_{SAM}, which is the estimator, based on respondent data, of $V_{SAM} = N^2(1/n - 1/N)S^2_{yU}$. Two things are noted: First, if the relationship is strong between the auxiliary x_k and the criterion y_k, the residuals e_k are small, and (4.2) shows that reweighting may cause only a modest addition to V_{SAM}. Secondly, the presence of the auxiliary variable x in the reweighted estimator helps to protect against bias for the event that some response model other than the assumed RHG model is really true (although the formulas just presented do not explicitly show this).

We do not here go into further detail about reweighting procedures; the reader may readily consult many existing references; a few are given at the end of this paper.

5. Imputation

Imputation is extensively used in surveys. It is interesting to note what Pritzker, Ogus and Hansen (1965) said 30 years ago about imputation policy at the US Bureau of the Census: "Basically our philosophy in connection with the problem of ... imputation is that we should get information by direct measurement on a very high proportion of the aggregates to be tabulated, with sufficient control on quality that almost any reasonable rule for ... imputation will yield substantially the same results ... With respect to imputation in censuses and sample surveys we have adopted a standard that says we have a low level of imputation, of the order of 1 or 2 percent, as a goal."

Today most surveys show much higher missing value rates than the desirable one or two percent. But imputation probably evolved into the widely used tool that it is today from the notion that inserting "a few artifical values" will not make much difference one way or the other. While this is true for one or two percent missing, it is not true for 30 to 40 percent missing, as is often the case today, even in surveys carried out by large government survey organizations.

Clearly, a key question is the quality of the imputed values. Are they close substitutes for the values missing? Statisticians well familiar with the subject matter of a specific survey can in many cases provide excellent imputations. The end result can in some cases be better than what would be obtained with reweighting.

However, the effects of imputation must not be ignored. Bias as well as increased variance will normally result. It is necessary to a) eliminate as far as possible the imputation bias and b) to explicitly provide users with the imputation variance, that is, to publish a value of the additional component of variance incurred by the imputation.

There is an extensive literature on imputation. Many imputation methods have been proposed. It is not the intention in this paper to review them. Four methods commonly used at Statistics Canada are mentioned in Section 7.

An basic distinction is the one between single value imputation and multiple imputation. In single value imputation, a single imputation is created for each missing value. A complete data matrix is obtained, in which the imputed values are flagged. Point estimates and variance estimates can be calculated with the aid of the completed set. A number of papers in recent years deal with this aspect.

In multiple imputation, two or more values are imputed for each missing value. Several completed data sets are thus obtained. Point estimates and variance estimates are calculated with the aid of the completed data sets. Multiple imputation was suggested by D. B. Rubin around 1977. His ideas are explained in a number of papers and in a book, Rubin (1987). Multiple

imputation effectively communicates the idea that imputation implies in itself a variability (in addition to the sampling variability). The variability within and between the several completed data sets is used in multiple imputation theory.

Rubin (1986) sees as a disadvantage of single imputation that "... the one imputed value cannot in itself represent uncertainty about which value to impute: If one value were really adequate, then that value was never missing. Hence, analyses that treat imputed values just like observed values generally systematically underestimate uncertainty, even assuming the precise reasons for nonresponse are known." It is certainly true that analyses that treat imputed values just like observed ones are misleading. It is not true that by imputing only one single value, one cannot give adequate expression to the total variance. In recent years, a number of papers propose variance estimates when single value imputation has been carried out. For a recent discussion of issue of the single versus multiple imputation, the reader is referred to Fay (1994), Kott (1995), Steel and Fay (1995).

6. Imputation at Statistics Canada

As in many other statistical agencies, imputation is widely used at Statistics Canada. Systems have been developed for computerized edit and imputation.

For the Canadian Census, a minimum change imputation methodology was formulated by Fellegi and Holt (1976). CANEDIT is a computerized edit and imputation system based on this methodology which has been used to impute qualitative variables in the Canadian census. A new minimum change hot deck imputation system called NIM (for New Imputation Methodology) has recently been developed, see Bankier, Fillion, Luc and Nadeau (1994). It is scheduled to be used in the 1996 Canadian census to carry out imputation for basic demographic variables. NIM works on the principle of nearest neighbour (or matching fields) donor imputation, and an advantage compared to CANEDIT is that NIM carries out imputation of qualitative and numeric variables simultaneously. It is claimed to be less inclined to falsely inflate the size of small but important population groups.

For Statistics Canada business surveys, the Generalized Edit and Imputation System (GEIS) is widely used to impute numeric variables. A description of GEIS is given in Cotton (1991). The statistician who uses GEIS can select from several imputation methods. The original development of GEIS focused on nearest neighbour imputation; other imputation methods were subsequently added to the system. At Statistics Canada, an imputation for a given survey may be elaborated by a team of methodologists and subject matter specialists. Once functional, the imputation for the survey is operated on a routine basis by the subject matter specialists.

Generalized systems play an important role in survey processing at Statistics Canada. For imputation, the Generalized Edit and Imputation System (GEIS) can be used to obtain a completed data set. For estimation, another of the generalized systems, the Generalized Estimation System (GES), can be used. GES is a micro-computer estimation package for survey data that has been recently been developed at Statistics Canada. It calculates the point estimates and the corresponding variance estimates for the domains of interest in the survey. The system is based on the GREG estimator, which is a general formulation for a wide range of specific estimators.

A typical scenario for edit, imputation and estimation in a Statistics Canada survey may thus be as follows: Following imputation using GEIS, the completed data set enters the GES, where estimates are computed. Each unit will be equipped with a weight w_k of the form $w_k = a_k g_k$, where $a_k = 1/\pi_k$ is the sampling weight and g_k is the g-weight expressed with the aid of the auxiliary information. Estimates are then easily obtained by summing the weighted values $w_k y_{*k}$ over each domain of interest in the survey. (A domain may consist of the entire population or any specified subpopulation.)

So far both the point estimation and the variance estimation in the GES has treated imputed values as if they are real observations. Since this is likely to lead to underestimation of the true variance, development is under way to provide more correct answers. In the near future, GES will be able to provide proper variance estimates when the survey data set have been completed by single-value imputation according to one (or several) of the most common methods: nearest neighbour imputation, ratio imputation, hot deck imputation, respondent mean imputation. The sampling variance component as well as an imputation variance component will be provided, since users of the GES have been expressing an interest in knowing how much of the total variance that derives from the imputation step.

7. Imputation Theory

Suppose single value imputation is used so that for a unit $k \in o$, the missing y_k is imputed by a value denoted $\hat{y}_k$. The resulting *y-data set after imputation* is then given by $\{y_{*k};\ k \in s\}$, where

$$y_{*k} = \begin{cases} y_k \text{ if } k \in r \\ \hat{y}_k \text{ if } k \in o \end{cases} \qquad (7.1)$$

It is also called the *completed data set* because a value is now present for every $k \in s$. Imputed values should be flagged for identification in the data

file, with a specification of the method used to impute. Several methods may be in use for one and the same data set.

As before, suppose that the survey design specifies a weight system $\{w_k\}$ to be used in the event of full response, so that Y would be estimated as $\hat{Y}_s = \Sigma_s w_k y_k$. The surrogate estimator (or the imputed estimator) is obtained by applying the specified weight system $\{w_k\}$ to the data after imputation:

$$\hat{Y}_{*s} = \Sigma_s w_k y_{*k} = \Sigma_r w_k y_k + \Sigma_o w_k \hat{y}_k \qquad (7.2)$$

where $o = s - r$. Note that $\hat{Y}_{*s}$ reduces to the full response estimator, $\hat{Y}_s = \Sigma_s w_k y_k$, in two cases: (i) when there is no nonresponse, that is, $r = s$; (ii) when the imputations are perfect substitutes so that $\hat{y}_k = y_k$ for every $k \in o$. Here, $\hat{Y}_s - Y$ is the sampling error, and the surrogate error $\hat{Y}_{*s} - \hat{Y}_s$ is in this case called the imputation error. The MSE decomposition (3.2) applies. In this case it can be written as

$$\mathrm{MSE}_{pq}(\hat{Y}_{*s}) = \mathrm{V}_{\mathrm{SAM}} + \mathrm{V}_{\mathrm{IMP}} + 2\,\mathrm{Cov}_p(\hat{Y}_s, \mathrm{B}_{\mathrm{IMP}|s}) + \mathrm{E}_p(\mathrm{B}^2_{\mathrm{IMP}|s}) \qquad (7.3)$$

where $\mathrm{V}_{\mathrm{SAM}} = \mathrm{V}_p(\hat{Y}_s)$ (the sampling variance), $\mathrm{B}_{\mathrm{IMP}|s} = \mathrm{E}_q(\hat{Y}_{*s} - \hat{Y}_s \,|\, s)$ (the conditional imputation bias), $\mathrm{V}_{\mathrm{IMP}|s} = \mathrm{V}_q(\hat{Y}_{*s} - \hat{Y}_s \,|\, s)$ (the conditional imputation variance), and $\mathrm{V}_{\mathrm{IMP}} = \mathrm{E}_p\{\mathrm{V}_{\mathrm{IMP}|s}\}$ (the unconditional imputation variance).

The analyst will often derive the imputed values via a prediction model using available auxiliary variables as predictors. Approaches of this kind are found in Hinde and Chambers (1991), Westat (1993), Deville and Särndal (1994). For our discussion, we refer to four imputation methods: Respondent Mean (RM) imputation, Hot Deck (HD) imputation, Ratio (RA) imputation, and Nearest Neighbour (NN) imputation. RA and NN imputation require auxiliary data: we assume that a value $x_k > 0$ is specified for every unit $k \in s$. The imputed value $\hat{y}_k$ for a unit $k \in o$ is given as follows in the four methods: *RM imputation*: $\hat{y}_k = \bar{y}_r = (1/m)\Sigma_r y_k$ for all $k \in o$, *HD imputation*: $\hat{y}_k = y_{\ell(k)}$, where $y_{\ell(k)}$ is the y-value of the donor unit, $\ell(k)$, drawn at random (with replacement) from the repondents $\ell \in r$; *RA imputation*: $\hat{y}_k = x_k \hat{\mathrm{B}}_r$, where $\hat{\mathrm{B}}_r = \bar{y}_r/\bar{x}_r$; *NN imputation*: $\hat{y}_k = y_{\ell(k)}$, where $y_{\ell(k)}$ is the y-value of the donor unit, $\ell(k)$, which is the unit closest to k in the sense that the minimum of $|x_\ell - x_k|$ over all potential donors $\ell \in r$ occurs for $\ell = \ell(k)$.

In the methods mentioned, the whole sample is taken as the imputation group. Alternatively, imputation can be carried out within subgroups of the sample. For example, in imputation by respondent subgroup mean, the imputed value is $\hat{y}_k = \bar{y}_{r_g} = (1/m_g)\Sigma_{r_g} y_k$ for all units in the nonresponse subgroup $s_g - r_g$; the subgroup identifier constitutes additional auxiliary information in this case. Another possibility is ratio imputation with a residual added to better preserve the form of the y-value distribution, $\hat{y}_k =$

$x_k\hat{B}_r + e^*_k$, where e^*_k is the result of selecting at random one residual from a given set such as $\{e_k = y_k - x_k\hat{B}_r: k \in r\}$.

8. Conditional Imputation Bias and Restoring Estimator

The conditional imputation bias is an important concept. If $B_{IMP|s} = E_q(\hat{Y}_{*s} - \hat{Y}_s \mid s) \approx 0$ whatever s, then the imputed estimator $\hat{Y}_{*s}$ agrees on average with the full response estimator $\hat{Y}_s$. In other words, the imputed estimator restores (on the average) the full response estimator. (Rancourt, Lee and Särndal (1994) call $\hat{Y}_{*s}$ a restoring estimator if $B_{IMP|s} \approx 0$ holds for any s.) We illustrate this property by some examples.

Example 8.1. Restoring property of RA imputation. Suppose the sample s is drawn with SRSWOR with the sampling fraction $f = n/N$ and that RA imputation is used as specified in Section 7: $\hat{y}_k = x_k\hat{B}_r$ with $\hat{B}_r = \bar{y}_r/\bar{x}_r$. The imputation error is

$$Z_r = \hat{Y}_{*s} - \hat{Y}_s = (N/n)\Sigma_o(x_k\hat{B}_r - y_k) = N\bar{x}_s E_r/\bar{x}_r \qquad (8.1)$$

where $E_k = y_k - x_k\hat{B}_s$ with $\hat{B}_s = \bar{y}_s/\bar{x}_s$. (Recall that overbar indicates arithmetic mean over the set given as an index, that is, $\bar{x}_s = \Sigma_s x_k/n$, $E_r = \Sigma_r E_k/m$, etc.) Note that given s, $\bar{x}_s E_r/\bar{x}_r$ has the form of a ratio estimator in the residuals E_k with x_k as the auxiliary variable. Given s, E_k is a constant value (independent of r). Can we conclude that $\hat{Y}_{*s}$ is a restoring estimator? The answer depends on the reponse mechanism. If it can be assumed to be uniform, then RA imputation is unbiased (and the imputed estimator is a restoring estimator of Y). To see this, condition on the number of respondents denoted m (which is random), and let E_m denote expectation with respect to the distribution of m. Given m, r behaves as an SRSWOR selection of m units from n. Using basic results for the ratio estimator, $E_q(Z_r \mid s,m) \approx (N/n)\Sigma_s E_k \approx 0$, which gives $B_{IMP|s} = E_m E_q(Z_r \mid s,m) \approx 0$. Note that the restoring property follows from an assumption about the response mechanism, without any reference to the linear regression model associated with ratio imputation. By contrast, under mechanisms other than the uniform, RA imputation can be considerably biased.

Example 8.2. Restoring property of RM imputation. RM imputation is the special case of RA imputation obtained when $x_k = 1$ for all k. It follows from the preceding example that RM is unbiased under the uniform mechanism (and the imputed estimator is restoring), but is otherwise likely to be biased.

Example 8.3. Restoring property of NN imputation. For NN imputation, it is also possible to show under certain assumptions that $B_{IMP|s} \approx 0$ under

the uniform mechanism. Under these conditions NN imputation is also approximately unbiased (and the imputed estimator is restoring).

Cases where $B_{IMP|s} \neq 0$ arise, for example, in estimation for domains; see Section 10.

9. Does Imputation Add Information?

The imputed values $\hat{y}_k$ in the RM and HD methods reutilise the y_k-values of the respondents. These imputation methods realize the goal of a complete data matrix, but no further information is added. The sample size "increase" from m to n is illusionary: the reasoning that "larger sample translates as smaller variance" is without foundation in this case.

By contrast, in using NN or RA imputation based on a strong auxiliary variable, appreciable gains of precision may be realized compared to RM and HD imputation. We illustrate this in the following two examples.

Example 9.1. Estimated variance components for RA imputation. Let the conditions be as in Example 8.1: The sample s of size n is drawn by SRSWOR with $f = n/N$; RA imputation is used so that $\hat{y}_k = x_k\hat{B}_r$ with $\hat{B}_r = \bar{y}_r/\bar{x}_r$. The response mechanism is assumed uniform. The imputation error $Z_r = \hat{Y}_s - Y_s$ is given by (8.1). Using known results for the ratio estimator it follows that

$$V_q(Z_r \mid s,m) \approx N^2(1/m - 1/n)(\Sigma_s E^2{}_k)/(n-1) \qquad (9.1)$$

so the conditional imputation variance is
$V_{IMP|s} = E_m V_q(Z_r \mid s,m) + V_m E_q(Z_r \mid s,m) \approx N^2\{E_m(1/m) - 1/n\}(\Sigma_s E^2{}_k)/(n-1)$.
The unconditional imputation variance $V_{IMP} = E_p(V_{IMP|s})$ is then estimated by

$$\hat{V}_{IMP} = N^2(1/m - 1/n)(\Sigma_r e^2{}_k)/(m-1) \qquad (9.2)$$

where $e_k = y_k - x_k\hat{B}_r$ with $\hat{B}_r = \bar{y}_r/\bar{x}_r$. Turning to the sampling variance V_{SAM}, an obvious estimator is

$$\hat{V}_{SAM} = N^2(1/n - 1/N)S^2{}_{yr} \qquad (9.3)$$

where $S^2{}_{yr} = \Sigma_r(y_k - \bar{y}_r)^2/(m-1)$. A somewhat better (reduced variance) estimator of V_{SAM}, derived via an argument in Rao and Sitter (1995), is

$$\hat{V}_{SAM} = N^2(1/n - 1/N)\{S^2{}_{er} + 2\hat{B}_r S_{exr} + \hat{B}^2{}_r S^2{}_{xs}\} \qquad (9.4)$$

where $S^2{}_{er} = \Sigma_r e_k{}^2/(m-1)$, $S_{exr} = \Sigma_r e_k x_k/(m-1)$.

The resulting estimate the total variance is $\hat{V}_{TOT} = \hat{V}_{SAM} + \hat{V}_{IMP}$.

Example 9.2. RA imputationis better than RM imputation. Intuitively, RA imputation is better than RM imputation under the conditions of Example 9.1 This superiority of RA imputation is revealed by comparing the imputation variance components in the two methods. For RM imputation, we get, setting $x_k = 1$ for all k in (9.2),

$$\hat{V}_{IMP} = N^2(1/m - 1/n)S^2_{yr} \qquad (9.5)$$

For RA imputation, $\hat{V}_{IMP}$ was given by (9.2), which is normally smaller than (9.5) if the relationship between x and y is strong (so that the residuals e_k are small). (Note that adding (9.3) and (9.5) we get the total estimated variance estimate under RM imputation as $\hat{V}_{TOT} = \hat{V}_{SAM} + \hat{V}_{IMP} = N^2(1/m - 1/N)S^2_{yr}$. It is reassuring to note that this represents the variance calculated for an SRSWOR sample of m from N.)

10. The Risk of Imputation Bias in Estimation for Domains

Estimation for a variety of domains (subpopulations) is required most surveys. In the presence of imputation, domain estimation requires special attention. One reason is that although the imputation method in use may give unbiased estimation for the whole population, bias may occur for any domain smaller than the whole population. This domain bias phenomenon is illustrated in the present section.

Let U_d be a typical domain; $U_d \subset U$. Its intersections with s, r and o are denoted $S_d = U_d \cap s$, $r_d = U_d \cap r$, and $o_d = U_d \cap o$. We want to estimate the y-total for the domain, $Y_d = \Sigma_{U_d} y_k$, which can be written as a total over the whole population, $Y_d = \Sigma_U y_k(d)$, by using the customary domain specific variable defined by $y_k(d) = y_k$ if $k \in U_d$, and $y_k(d) = 0$ otherwise.

Corresponding to current practice we assume that the imputation is carried out once and for all using the entire sample as the imputation group. That is, as described in Section 7, imputed values $\hat{y}_k$ are created for $k \in o$. The resulting completed data set (7.1) is used for computing estimates for the whole population or for a domain, with the aid of the prespecified weight system $\{w_k\}$. Thus the estimator of Y_d for full response would be $\hat{Y}_s(d) = \Sigma_s w_k y_k(d) = \Sigma_{s_d} w_k y_k$. The implied imputed estimator (the surrogate estimator of Y_d) is therefore

$$\hat{Y}_{*s}(d) = \Sigma_s w_k y_{*k}(d)$$

where $y_{*k}(d) = y_{*k}$ if $k \in s_d$ and $y_{*k}(d) = 0$ if $k \in s - s_d$. Equivalently, we can write

$$\hat{Y}_{*s}(d) = \Sigma_{r_d} w_k y_k + \Sigma_{o_d} w_k \hat{y}_k \qquad (10.1)$$

That (10.1) may be biased becomes apparent using the framework in Sections 2 and 3 More particularly, attention is focused on the conditional imputation bias, as in the following example.

Example 10.1. Bias of RA imputation when estimating for a domain smaller than the entire population. Let the conditions be as in Example 8.1: SRSWOR sampling with $w_k = N/n$ for all k, RA imputation with $\hat{y}_k = x_k \hat{B}_r$ with $\hat{B}_r = \bar{y}_r/\bar{x}_r$. Let us assume that the response mechanism is uniform. Denote as n_d and m_d the (random) sizes of s_d and r_d. To find out if (10.1) is a restoring estimator, we examine the imputation error $Z_r(d) = \hat{Y}_{*s}(d) - \hat{Y}_s(d)$. Conditioning on n_d and m_d we get

$$E_q(Z_r(d) \mid s, n_d, m_d) \approx (N/n)(n_d - m_d)\bar{x}_{s_d}(\hat{B}_s - \hat{B}_{s_d}) \qquad (10.2)$$

where $\hat{B}_s = \bar{y}_s/\bar{x}_s$, $\hat{B}_{s_d} = \bar{y}_{s_d}/\bar{x}_{s_d}$. The conditional bias, given s, is the expected value of (10.2) with respect to the respective distributions of n_d and m_d. But it is clear already from (10.2) that the imputed estimator (10.1) is conditionally biased whenever $\hat{B}_s \neq \hat{B}_{s_d}$, and the procedure is not restoring. A bias correction can be applied. We can, for example, correct (10.1) by subtracting an estimate of the bias (10.2), $(N/n)(n_d - m_d)\bar{x}_{o_d}(\hat{B}_r - \hat{B}_{r_d})$, where $\hat{B}_{r_d} = \bar{y}_{r_d}/\bar{x}_{r_d}$. The resulting bias corrected estimator of the domain total $Y(d)$ is

$$\hat{Y}_{*s}(d)_{corr} = (N/n)\{\Sigma_{r_d} y_k + \Sigma_o (x_k \hat{B}_{r_d})\}$$

This corrected form conveys the message that to obtain a restoring estimator of the domain total, we should have imputed $x_k \hat{B}_{r_d}$ rather than $\hat{y}_k = x_k \hat{B}_r$ for the missing value y_k. Note that the preferred imputations $x_k \hat{B}_{r_d}$ are domain dependent. In other words, the specific domain, not the entire sample, should have served as the imputation group. After a moment of reflection, this comes as no surprise. However, imputing differently for each domain of interest meets with practical difficulties. In a large survey, one will not often go to this trouble, especially if the domains of interest are numerous. Some imputation bias may have to be accepted.

References

Bankier, M., Fillion, J. M., Luc, M. and Nadeau, C. (1994). Imputing numeric and qualitative variables simultaneously. *Statistics Canada report*, September 7, 1994.

Bethlehem, J. G. (1988). Reduction of nonresponse bias through regression estimation. *Journal of Official Statistics*, 4, 251–260.

Binder, D. A. (1991). A framework for analyzing categorical survey data with nonresponse. *Journal of Official Statistics*, 7, 393–404.

Cotton, C. (1993). GEIS: Functional description of the generalized edit and imputation system. *Business Survey Methods Division, Statistics Canada*, July 25, 1993.

Deville, J.–C. and Särndal, C. E. (1994). Variance estimation for the regression imputed Horvitz–Thompson estimator. *Journal of Official Statistics*, 10, 381–394.

Ekholm, A. and Laaksonen, S. (1991). Weighting via response modeling in the Finnish household budget survey. *Journal of Official Statistics*, 7, 325–337.

Estevao, V., Hidiroglou, M. A. and Särndal, C. E. (1995). Methodological principles for a generalized estimation system at Statistics Canada. *Journal of Official Statistics*, 11, 181–204.

Fay, R. E. (1994). Analyzing imputed survey data sets with model assisted estimators. Proceedings, *Section on Survey Research Methods, American Statistical Association*, 900–907.

Fellegi, I. P. and Holt, D. (1976). A systematic approach to automatic edit and imputation. *Journal of the American Statistical Association*, 71, 17–35.

Hinde, R. L. and Chambers, R. L. (1991). Nonresponse imputation with multiple sources of nonresponse. *Journal of Official Statistics*, 7, 167–179.

Kott, P. S. (1994). A paradox of multiple imputation. *Presented at American Statistical Association annual meeting, Orlando, Florida.*

Lee, H., Rancourt, E. and Särndal, C. E. (1994). Experiments with variance estimation from survey data with imputed values. *Journal of Official Statistics*, 10, 231–243.

Pritzker, L., Ogus, J. and Hansen, M. H. (1965). Computer editing methods: some applications and results. *Bulletin of the International Statistical Institute*, 41, 442–466.

Rancourt, E., Lee, H. and Särndal, C. E. (1994). Bias corrections for survey estimates from data with ratio imputed values for confounded nonresponse. *Survey Methodology*, 20, 137–147.

Rao, J. N. K. and Sitter, R. R. (1995). Variance estimation under two-phase sampling with application to imputation for missing data. *Biometrika*, 82, 453–460.

Rubin, D. B. (1986). Basic ideas of multiple imputation for nonresponse. *Survey Methodology*, 12, 37–47.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley.

Särndal, C. E. and Swensson, B. (1987). A general view of estimation for two phases of selection with applications to two-phase sampling and nonresponse. *International Statistical Review*, 55, 279–294.

Särndal, C. E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling* New York: Springer-Verlag.

Steel, P. and Fay, R. E. (1995). Variance estimation for finite populations with imputed data. *Presented at American Statistical Association annual meeting, Orlando, Florida.*

Westat (1993). Imputation for item nonresponse in surveys. *Westat report*, October, 1993.

NONRESPONDENTS IN A MAIL SURVEY: WHO ARE THEY?

Lisette Bros, Edith de Leeuw, Joop Hox and Gerard Kurver
Department of Education, University of Amsterdam

Respondents and nonrespondents to a mail survey on educational and professional career were compared using multivariate analysis. Respondents tended to have a higher score on an intelligence test and do better in the first grades of primary school. They also more often come from a better social economical background.

Key words: Characteristics of nonrespondents, economic status, education, family background, gender, household size, intelligence, urbanization, mail survey, school career.

1. Introduction

If nonresponse were a random process and respondents and nonrespondents were interchangeable, survey nonresponse would not worry survey methodologists. Unfortunately, in general nonrespondents do differ from those who participate in a survey so there is a reason to worry about nonresponse error.

Studying the characteristics of respondents and nonrespondents has both a theoretical and a practical goal. Knowing how respondents differ from nonrespondents helps to formulate theories on reasons for nonresponse; it also provides information for statistical adjustment and nonresponse bias correction. However, collecting information about nonrespondents is not easy: a direct approach of nonrespondents has its own nonresponse problems, but may result in rich data; using information from other sources (e.g., sampling frame) avoids the low response rate of a survey of nonrespondents, but typically does not produce very rich data (Groves 1989, 186). We were in the fortunate circumstance to have access to a rich data base providing information on respondents and nonrespondents in a mail survey. In this contribution we report on our major findings analyzing this data. We start with an overview of the main results of studies into nonrespondent characteristics. We then move on to a description of our dataset, the methods used, and the main results. We end with a critical summary.

2. Background Variables and Respondent Characteristics Associated with NonResponse

Our main focus of interest is nonresponse in mail surveys; in a mail survey potentially everyone is reached, but some may decide not to cooperate. Therefore, in the next section only known correlates of non-cooperation (refusal) are discussed and characteristics of difficult-to-reach persons and not-at-homes are not included. Also it should be noted that most studies report distinct attributes of nonrespondents and fail to investigate multivariate relations (cf. Groves 1989, Goyder 1987). When results are confirmed in multivariate analyses this will be explicitly mentioned.

An easy and frequently studied variable is respondent's *gender*; the results however are inconclusive and no clear trend can be discerned. Sometimes a higher nonresponse among men is reported (Groves 1989, 202; Brehm 1993, 30), but the results are confounded by accessibility and do not hold in multivariate analysis (Goyder 86, 100).

Age of the respondent is a strong correlate of nonresponse in many countries, that is, elderly are more reluctant (Groves 1989, 202–204; Goyder 1987, 85; Couper and Groves 1993; Barnes 1990). This relationship holds when one controls for other variables in a multivariate analysis (Goyder 100).

Social Economic Status or *SES* has been found to correlate with nonresponse in mail, telephone and face to face (Goyder 1987, 83–84). However SES is a fuzzy concept, and indicators of SES may be education, occupation, income, or a combination of these variables. *Education* is a correlate of nonresponse in mail surveys, but also in face to face and telephone interviews. More highly educated tend to respond better (Groves 1989, 205–206; Foster and Bushnell 1993). *Occupation* does correlate with nonresponse with lower nonresponse among the higher status occupations (Barnes 1990); this relationship holds in a multivariate logistic regression Goyder (1987, 99). Finally, the results on *income* are inconsistent and it has been suggested that the relationship might be nonlinear (Couper and Groves 1993).

The findings for *household size* suggest that response increases with household size, especially one and two person households are associated with a low response (Groves 1992, 208; Couper and Groves 1993; Barnes 1990). However this finding could be confounded with the effects of age, as elderly disproportionally live alone.

Urbanicity is a strong correlate of nonresponse with lower response in the urban regions (Groves 1989, 233–234; Goyder 1987, 88; Foster and Bushnell 1993). Couper and Groves (1993) discuss three other aspects of living conditions: home ownership, property value and housing structure. Results tend to be inconclusive, but there is a tendency for lower response in *multi-unit structures* and a higher response from single family homes.

Independent of the nonresponse studies summarized above, interesting studies have been done in the field of psychology. Experimental psychologists have been concerned whether volunteers (respondents) in psychological experiments differed from nonvolunteers (nonrespondents) and are not representative of human subjects in general. Rosenthal and Rosnow (1975; chap. 3) summarize the findings on volunteer characteristics. Psychological research confirms the main results from survey research: volunteers tend to be better educated, have a higher occupational status, and tend to be younger. In addition, they report that volunteers tend to score higher than nonvolunteers on tests of *intelligence*. Also, volunteers tend to be more often *first born*. The evidence on *religious* orientation is inconclusive.

In our study we could test all the relations mentioned above, except the relationship of age and nonresponse, since all our potential respondents were born in the years 1957–58 and are therefore in the same age group. It is clear that many of the variables mentioned above are themselves intercorrelated. Therefore, in addition to bivariate analyses we performed a multivariate analysis.

3. Method and Dataset Used

In general, sampling frames do not contain much information, but we were in the unusual position to have access to a sampling frame that contained rich information. The sampling frame consisted of the case records of all pupils that were in the first grade of primary school in 1964–65 in the Dutch city of Enschede. In the early sixties a large educational panel study was conducted: the Enschede cohort (Van Calcar 1968; Bros in preparation). Available were data on family background, intelligence, and school performance (educational tests and repeating class).

To this database we added information linked the *present* addresses of the sampling units through their zipcodes. The Dutch zipcodes form an extremely fine grid, and detailed socio-economical background information is available on clusters of, on average, 15 households each. This zipcode based information is collected mainly from interviews by telephone with added information from registers and databases. It is typically used for commercial purposes such as direct mailing, but can also provide efficient information for nonresponse research.

In November 1992 a questionnaire on educational and professional career was sent out to all pupils still alive (N = 2265). The initial response was 34 %. Two weeks later a letter was sent out to remind the respondents of the questionnaire. The reminder raised the response to 49.6 %. In total, 1142 persons (50.4 %) did not respond; this number includes all refusals (7). The database analyzed here consists of both the data from the original individual based sampling frame (Enschede-cohort) and the added current zipcode based

information; the latter are indicated in the results section with '*zip*'. The multivariate modeling is done using logistic regression with the binary dependent outcome variable response; the dependent variable response was coded 1 for a completed, returned questionnaire and 0 for nonresponse including refusals.

4. Results

Bivariate Analyses

We start with an examination of the results of various bivariate analyses. There was a significant effect of gender ($p<.05$) on response, that is a higher nonresponse among men.

When we look at social and economic status, we see that family social status as indicated by the occupation of the father of the sampling units has a significant effect ($p<.01$): more nonrespondents come from low status families, and more respondents from middle status families. This is confirmed by the analysis on zipcode based income information of the respondent; groups in the higher income brackets tend to respond more often ($p<.01$). Finally education, as indicated by a successful school career in the primary school (not repeating classes), shows a significant effect ($p<.01$): more respondents never had to repeat a class. This is also supported by the zip-code based information; respondents more often belong to the higher educated ($p<.01$).

Household size (zip) has a small effect on response ($p<.05$), indicating a slight tendency for larger households to respond more often. Urbanicity (zip) shows a clear tendency ($p<.01$) with persons living in less urban regions responding more often. Whether or not respondents live in multi-unit dwellings (zip) has hardly any influence on response (marginally significant $p = .05$). But the age of the houses does matter: persons living in houses that were build fairly recently (i.e., after 1970) more often respond than those living in older neighborhoods. Other housing variables that are related to economic status (zipcode based: home-owner ship and value of property) show that whether one owns a house effects response positively ($p<.01$), but value of property does not significantly covary with response.

Going from sociological to psychological variables we find a strong effect of intelligence ($p<.01$): respondents score higher on the Primary Mental Ability test (Thurstone and Thurstone 1954). There also is a slight effect of sibling position as child ($p<.05$: respondents are in general the elder children in a family) and religious ties in youth ($p<.01$: more respondents went to a school with a religious denomination, especially Roman Catholic).

Multivariate Analysis

Although most indicators of nonresponse are themselves related, most nonresponse research ignores this and reports bivariate analyses only. Therefore, some of the reported relations with nonresponse may be spurious relationships and multivariate analysis is needed to disentangle these effects (cf. Groves 1989; Goyder 1987). We used logistic regression. As the predictors come from two different sources we started with two separate analyses: one with the historic individual educational variables as predictors, and the second with the address based zipcode variables as predictors. In the final analysis we used all predictors available. The results, summarized in Table 1, clearly confirm the point taken by Groves and Goyder that many relationships do not hold in a multivariate analysis. Only the educational variables repeating classes in primary school, score on intelligence test, school denomination and the address based variables urbanicity and age of house show a clear effect on survey response in the final analysis.

To judge the impact of a variable on response the regression weight does not provide enough information; one should also know the scale on which the variable is measured. The variable 'repeating a class' is measured on a 3-point scale: 0 (never), 1 (once), 2 (twice); 'urbanicity' on a 7-point scale (1 = very small settlements, 7 = very large cities); 'own house' is a dichotomy (0 = no, 1 = yes); 'age house' is an 8-point scale (1 = build after 1989, 2 = between '80–'89, 3 = '70–'79, ..., 7 = 1800–1900, 8 = older than 1800); score on IQ test is measured as a z-score (continuous, mean 0, s.d = 1); and religious ties on a 4-point scale (1 = no, 4 = strong).

Table 1. Effect of Respondent Characteristics on Response in Logistic regression Bivariate correlations, unstandardized regression coefficients and their standard errors. Only significant regression coefficients reported.

Independent variable	Bivariate correlation	Multivariate logistic regression		
		Historic	Zip	All variables
gender	.05	n.s.	—	n.s.
occupation father	.09	n.s.	—	n.s.
income (zip)	.15	—	n.s.	n.s.
repeating a class	-.22	-.58 (.14)	—	-.53 (.16)
education (zip)	.16	—	n.s.	n.s.
household size (zip)	.06	—	n.s.	n.s.
urbanicity (zip)	-.10	—	-.11 (.04)	-.14 (.04)
own house (zip)	.18	—	.18 (.05)	n.s.
value house (zip)	n.s.	—	n.s.	n.s.
age house (zip)	-.13	—	-.14 (03)	-.08 (.04)
multiunit (zip)	-.05	—	n.s.	n.s.
score on IQ-test	.24	.34 (.06)	—	.39 (.07)
sibling position	-.07	n.s.	—	n.s.
religious ties (denomination school)	.09	.10 (.04)	—	.09 (.04)

The regression coefficients in the logistic regression models are on the logit scale. For interpretation it is useful to consider what would happen to a basic response rate of 50 % if we go from the one extreme to the other by transforming the predicted value back to percentage points. Repeating a class has three scale points: 0, 1, and 2. Thus, having repeated classes twice lowers the expected basic response rate 24.3 points from 50 % to 25.7 %. Going from very small settlements (urbanicity = 1) to large cities (urbanicity = 7) lowers the expected response rate 19.8 points from 50 % to 30.2 %, and in a neighborhood with old houses (built before 1800) the response drops with 13.6 points to 36.4 %. Having come from a school with strong religious ties leads to a rise of 6.7 points in the expected response rate. Finally, IQ (as measured in childhood) has a strong effect. Going from one standard deviation below the mean to one standard deviation above the mean (about 67 % of the population falls between these limits) raises the expected response by 18.6 percentage points from 50 % to 68.6 %.

5. Summary and Discussion

The results of the multivariate analyses did indicate that both individual attributes and neighborhood characteristics influence nonresponse. The importance of multivariate nonresponse analysis is clearly illustrated; for instance when controlled for the effect of other variables, no clear gender effect could be found. Two education related variables – intelligence tested in primary school and having to repeat a class – have a strong influence on nonresponse. Ownership of house is the only clear economical variable that showed a significant effect. The often reported finding that urbanicity is a strong correlate of nonresponse was confirmed. Another neighborhood related variable that showed to be important is the age of the house, confirming the anecdotic testimony of (Dutch) interviewers that in certain 19-century neighborhoods it is difficult to get a high response. Finally, a religious background as indicated by school denomination, does have a slight positive effect on response even when we controlled for other variables.

When these indicators were incorporated in a logistic regression, no significant effects could be detected for indicators that are sometimes named in the literature on nonresponse, such as gender, income, and living in multi-unit structures. However, since we investigated a specific school cohort, we have a very restricted age range. It is possible that certain variables in Table 1 would have a larger effect if this restriction of range was not present.

The data were derived from a mail survey on educational and professional career of Dutch young-middle aged adults (± 35 years old). The results show that nonresponse did NOT occur at random. Respondents and nonrespondents did not just differ, but they also differed on variables such as intelligence and repeating classes that could be strongly connected with the theme of the questionnaire. When reporting the findings of this specific survey, nonresponse adjustment based on a model derived from Table 1 is advised.

Acknowledgments: We thank Geo-marktprofiel for their permission to use their zip-code information for this study.

References

Barnes, B. (1990). Studies of nonrespondents in British government household surveys. *Stockholm 1990 workshop on household survey nonresponse.*

Brehm, J. (1993). The phantom respondents; Opinion surveys and political representation. *Ann Arbor, The University of Michigan Press.*

Bros, L. Als ik groot ben, dan... [Research into the educational and professional career of the generation that went to school in 1964–65 in Enschede]. Ph.D. Dissertation (in preparation).

Couper, M. P. and Groves, R. M. (1993). Household level effects on survey participation. *Bath 1993 workshop on household survey nonresponse.*

Foster, K. and Bushnell, D. (1993). Nonresponse bias on government surveys in Great Britain. *Ottawa 1994 workshop on household survey nonresponse.*

Goyder, J. (1987). *The silent minority; Nonrespondents on sample survey.* Cambridge: Polity Press.

Groves, R. M. (1989). *Survey errors and survey costs.* New York: Wiley.

Thurstone, T. G. and Thurstone, L. L. (1954). Examiner manual for the SRA Primary Mental Abilities for ages 5–7. *Chicago: Science Research Association,* Inc. 4th Edition.

Van Calcar, C. (1968). Onderzoek en begeleiding in het basisonderwijs in Enschede [Research and coaching in primary education in Enschede]. *Enschede: Pedagogisch Centrum.*

A NEW PROGRAMME OF NONRESPONSE RESEARCH AT SCPR, LONDON

Pamela Campanelli
Survey Methods Centre,
Social and Community Planning Research
United Kingdom

In July of 1995, Social and Community Planning Research (SCPR) began a programme of research funded by a grant from the UK Economic and Social Research Council (ESRC) to take an in-depth look at the role of interviewers in the survey nonresponse process. This 18 month long programme is in co-operation with the ESRC Centre on Micro-Social Change at the University of Essex and NOP Research. It focuses on three separate areas. The first involves isolating the effects of interviewers on survey nonresponse from other factors such as the characteristics of the area where the interviewer is working and the particular characteristics of the address/respondent. This will be based on a specially designed experiment and will make use of a multilevel modelling approach to the cross-classification: interviewers by areas. The second and third involve new ways of investigating and analysing interviewers' patterns of making contact with resident(s) of each sample address and persuading the resident(s) to take part in the survey. This paper describes our research plans.

Key words: Nonresponse, surveys, interviewers, doorstep interaction, multilevel modelling.

1. Introduction

A serious hazard to survey research is that a substantial proportion of the representative sample of the population originally selected may fail to respond. There are very broadly two strategies for addressing the problem of nonresponse bias (if one rejects what is in fact the most common strategy in practice – that of simply ignoring it). The first is to accept that nonresponse is a fact of survey life and try to mitigate its biasing effects by post-hoc weighting. The second and, if feasible, clearly the preferable strategy is to reduce the proportion of nonresponse which occurs in the first place. The experience of survey agencies which have succeeded in maintaining response shows that strategies such as increasing the number of calls which

interviewers are required to make before abandoning an address, lengthening the fieldwork period, sending advance letters to sample addresses and offering incentives to respondents can all make a contribution, though they all involve some extra cost in time and money.

As face-to-face field interviewing is a key method of data collection on social and market research surveys in the UK and is likely to remain so for the foreseeable future, there is a strong case for looking at the role of the survey interviewer in the nonresponse process. For example, past research has shown how critical the initial doorstep interaction between interviewer and address residents, typically lasting less than a minute, is in determining whether or not the prospective respondent will agree to cooperate (e.g. Morton-Williams 1993; Groves and Cialdini 1991). Past research has also suggested that the efficiency with which interviewers call and recall on addresses affects not only the cost per interview, but also the proportion of addresses which ultimately yield successful interviews (e.g. Lievesley 1986).

Our programme of research involves three empirical studies to examine in detail the role that interviewers play in the nonresponse process, building on the work of Morton-Williams (1993), Lievesley (1986), and Groves and Cialdini (1991). Each of these will be described in turn.[1]

2. Sub-project 1: Separating Interviewer from Area and Respondent Effects on Nonresponse

The aim of the first sub-project is to tease out the interviewer effect on nonresponse from other effects. This is far from straightforward, since interviewers in the UK are normally assigned clusters of addresses within particular areas and the topographic, demographic, socio-economic and other attributes of areas and their inhabitants are themselves known to influence the average probability of securing a successful interview as a result of a survey approach. In order to separately estimate these various effects one needs an `interpenetrated' design in which interviewers are assigned at random to areas and there are at least two interviewers per area. Due to field cost considerations, such designs are rare.

This sub-project makes use of data from the second wave of the BHPS which contains a modified interpenetrated design in a subset of areas. The interpenetrated design was set up by Pam Campanelli (while she was a member of the BHPS survey team) and Colm O'Muircheartaigh (from the LSE). The BHPS is an annual panel survey which began in 1991 covering all of Great Britain involving approximately 5,000 households. A timely

[1] We had originally proposed a fourth study to develop and test improved interviewer training modules. This component was not funded.

advantage to this sub-project is the recent availability of multilevel software which will handle cross-classified models (see Goldstein 1995; Rasbash et al. 1995). This software allows one to take advantage of the hierarchical nature of the data while being able to separate area from interviewer effects. A multilevel cross-classified approach to the analysis will tell us what factor is having more of an effect on nonresponse: areas or interviewers. This type of model has several advantages over the standard ANOVA approach with the primary one being the ability to include interviewer and area characteristics in the model as explanatory variables for these sources of variability. An advantage of using the BHPS data is that the interpenetrated sample design was implemented in Wave II of the panel study. Thus the characteristics of Wave II nonrespondents as well as respondents are known and can be added to the modelling process. Limited information on Wave I respondents is also available.

Analyses will be conducted separately for interviewers' refusal and non-contact rates, as these will require separate solutions. Analysis of interviewer characteristics will have direct implications for interviewer recruitment practice. This sub-project will identify where training needs to be focused. The next two sub-projects will expand this knowledge and provide clues on how to accomplish this.

3. Sub-project 2: Focus on Interviewers' Calling Strategies

General survey knowledge suggests that the timing of an interviewer's call on a household affects the success of that call. People are more likely to be home at certain times of days than others and this affects the non-contact component of nonresponse. Thus interviewers are typically advised to vary the time of day and day of the week on which they call in order find people at home. Implicit in fieldwork strategies developed through intuition and experience is the idea that time of day affects the refusal component of nonresponse (e.g., calling times are guided by social conventions excluding early morning and late evening calls).

Such survey advice has mainly been gathered through the general experience of fieldwork agencies. There are a few published examples, e.g. Weber and Burt (1972) and Weeks et al. (1980) in the USA and Swires-Hennessy and Drake (1992) and Lievesley (1986) in the UK. These studies, with the exception of Lievesley (1986), have mainly focused on interviewers' first calls at addresses in order to identify the probabilities of finding someone at home at particular times.[2] Data on hours of `wakeful occupancy'

[2] There is also literature on optimising the call strategies of telephone interviewers. See for example, Groves and Robinson (1982).

are also available from time use studies (cf Hill 1978). More information on when respondents are approachable would help to maximize the efficiency of interviewers' calls and has direct implications for survey field costs. Information on the way interviewers organise their interviewing time is as important. Lievesley (1986) found that interviewers who had lower non-contact rates were not calling more frequently, but more effectively.

We propose the use of three datasets for this sub-project. These include the BHPS which is collected by NOP Research, the portion of the Family Resources Survey (FRS) which is conducted by SCPR, and SCPR's British Social Attitudes (BSA) surveys. The FRS is a continuous survey of households on detailed financial matters conducted jointly by SCPR & OPCS, entailing interviews with approximately 25,000 households every year. The BSA is an annual survey conducted by SCPR which began in 1983. The 1994 design yielded a sample of 3,600 individuals from across Great Britain. These three datasets will allow us to observe contrasts between organisations and between different types of call processes, i.e. attempting to interview everyone in a household (BHPS and FRS) and interviewing a randomly selected adult (BSA).

As with Sub-project 1, we will be bringing in data on interviewer, respondent and Census area characteristics, where applicable and available. Basic analyses will look at the time of day and day of the week of the first call. Subsequent analyses will look at the call record as a whole to find the factors which best predict the final success or failure. Following on the work of Groves and Cialdini (1991), we expect there to be a complex web of factors where the nature of the interaction of an earlier contact has independent effects on the outcome of the next contact. Some of the analyses planned may best be conceptualised in an event history framework (Blossfeld, Hamerle and Mayer 1989).

Through a interviewer questionnaire and examination of the call-records, this sub-project will provide information as to what contact and location strategies interviewers use and what components of these strategies have the largest effect on the survey non-contact rate and survey cost. Efficient components can then be developed into generalisable training schemes.

4. Sub-project 3: Focus on Interviews' Persuasion Strategies

Groves and Cialdini (1991) suggest that individuals typically use varying degrees of two information processing strategies, referred to as `systematic' and `heuristic', in making decisions. They go on to suggest that a systematic strategy is based on a rational assessment and depends on the individual's interest, time, energy, and cognitive capacity; however, when someone is distracted, tired, or indifferent, a 'heuristic' strategy becomes increasingly probable, leading individuals to make their decisions on past behaviour in

similar situations. It is likely that the later strategy applies predominantly to survey respondents. From research in social psychology on compliance with requests, helping tendencies, and persuasive appeals, Groves and Cialdini (1991) have developed a theory of survey participation which we intend to explore using the methodology developed by Morton-Williams (1993). Details about the proposed sub-project implementation are included in Annex 1.

Through the detailed examination of the complete doorstep exchange, this sub-project will allow us to see if Morton-Williams (1993) findings are replicated (e.g., she suggests that the initial component of a good introduction is short, plain, and economical) and whether the better interviewers make more use of strategies tailored to the specific respondent as the Groves and Cialdini's theory of participation would predict. As well as being of interest to survey methodologist, this information will give a comparison of the strategies interviewers think they use and what they actually use. In addition to informing training, Groves and Cialdini (1991) suggest that documenting key features of the interaction between interviewers and respondents may in time permit the estimation of a general propensity to respond, which could be employed for post-survey adjustment procedures which reduce nonresponse bias.

References

Blossfeld, H., Hamerle, A. and Mayer, K. (1989). *Event History Analysis.* Hillsdale, NJ: Lawrence Erlbaum.

Goldstein, H. (1995). *Multilevel Statistical Models* (second edition). London: Edward Arnold; New York: Halstead Press.

Groves, R. and Cialdini, R. (1991). Toward a Useful Theory of Survey Participation. *Proceeding of the Section on Survey Research Methods, Joint Statistical Meetings. Alexandria, VA: American Statistical Association.*

Groves, R. and Robinson, D. (1982). Final Report on Callback Algorithms on CATI Systems. *Washington, DC: report to the U.S. Bureau of the Census.*

Hill, D. H. (1978). Home Production and the Residential Electric Load Curve. *Resources and Energy*, 1, 339–358.

Lievesley, D. (1986). Unit Nonresponse in Interview Surveys, JCSM unpublished paper. London: SCPR.

Morton-Williams, J. (1993). *Interviewer Approaches.* Aldershot: Dartmouth Publishing Company Limited.

Rasbash, J., Woodhouse, G., Goldstein, H., Yang, M., Howarth, J. and Plewis, I. (1995). *MLm Software.* London: Multilevel Models Project, Institute of Education, University of London.

Swires-Hennessy, E. and Drake, M. (1992). The optimum time at which to conduct interviews, *Journal of the Market Research Society*, 34(1).

Weber, D. and Burt, R. C. (1972). Who's Home When. *Washington, D.C.: US Bureau of the Census.*

Weeks, M. F., Jones, B. L., Folsom, R. E. and Benrud, C. H. (1980). Optimal Times to Contact Sample Households. *Public Opinion Quarterly*, 44(1), 101–114.

Annex 1

ESRC Project: Interviewers and Survey Nonresponse
Proposal for Collection of Doorstep Introduction Information

Thirty-two interviewers are to participate:

16 NOP (former BHPS) interviewers now doing Political Tracking	— 1 respondent per household
16 SCPR (FRS) interviewers	— Whole household interviewed

Interviewers to be selected to allow for geographic spread (but probably not including Scotland) so as to minimise travel and hotel costs. Also selected to allow for a range of experience levels.

The experiment will apply to all households in an interviewer's workload. The average workload size for Political Tracking is 32 addresses. The average workload size for FRS is 24 addresses. Assuming the loss of 12 percent due to deadwood and an increase of 2 percent due to multiple households at an address, this yields doorstep information on 778 different households. Not all of these cases, however, will be equally useful for analysis. The main interest will be in those interactions where the interviewer has had a chance to talk with the respondent and been required to answer respondent questions or used his/her persuasive skills.

As the same tape recorders will be used for each piece of work, the two field periods must not overlap. The current schedule is as follows:

NOP interviews	— December '95
SCPR interviews	— February/March '96

As one of the goals of the project is to investigate the feasibility of using an interviewer debriefing form to capture essential elements of the doorstep interaction, 2 main methods of collecting doorstep information will be used and experimentally compared. Each interviewer will implement each method in a randomly assigned portion of the households. A third method will be used in a small subset of cases. The 3 methods are as follows:

Method 1	— Interviewer Form
Method 2	— Interviewer Form + Taping
Method 3	— Interviewer Form + Observer

Note that:

- Methods 1 and 2 are to be used for <u>all</u> calls at a given address and over all addresses. If a given call is a non-contact, this can simply be recorded as such. One tape will be used for each address. (Jean Morton-William's design was similar to this. This will allow us to have information on the whole call-record of calls to study how the interviewer changes or does not change tactics from call to call. Re-issues will not be examined.)
- Method 3 will be identical to Method 1 with the addition of the placement of an observer on a random selection of calls.

There is also the possibility of adding some confederate respondent cases to the workload. This will be difficult to implement in the FRS as there is a long precedent of interviewers only receiving a 24 address workload and as the CAPI sample files and procedures do not lend themselves easily to the addition of special cases.

The interviewer forms to be based on forms developed by Groves and Couper, plus

- Some additional measures of non-verbal behaviour
- An indicator of respondent reluctance
- Basic demographics and housing characteristics on all respondents and nonrespondents

At the analysis stage we will be:

- Comparing collection methodologies
- Looking for evidence of tailoring
- Looking at pairs of behaviours
- Looking for evidence of interviewer retreat and re-advance to avoid refusal
- Etc.

AN EXPLORATION OF NONRESPONSE IN THE 1990 NATIONAL ELECTION STUDY

Mick P. Couper and Robert M. Groves
Survey Research Center, University of Michigan
and Joint Program in Survey Methodology

We use a set of contact descriptions questions from the 1990 National Election Studies to explore the nonresponse implications of the householder-interviewer interactions at the time of the survey request. For nonresponse adjustment to be effective in reducing nonresponse bias, the variables used in the adjustment must be related to both propensity to cooperate and to the key substantive variables. We explore correlates of survey cooperation and potential impacts of nonresponse error, using adjustments based on the contact description data.

Key words: Nonresponse bias, nonresponse adjustment, householder-interviewer interaction.

1. Introduction

In many surveys, response rates are used as a proxy of quality; however, nonresponse error is function of both the nonresponse rate and the difference between respondents and nonrespondents. The task of the nonresponse adjustor is one of predicting both response propensity and the survey variables jointly. As Elliot (1991) notes, three conditions must hold for nonresponse weighting to reduce bias: (1) the response rates must be different in the different weighting classes, (2) the means for major survey variables must be different in different classes, and (3) within weighting classes, the means of the major survey variables must be similar for respondents and nonrespondents. While the third assumption is generally untestable, the first two can be evaluated using the survey data. The focus of most *post hoc* nonresponse adjustment appears to be on the first of these conditions, making use of variables available on the sample frame (e.g., urbanicity, region) that are associated with differential response propensities.

A common approach in the literature to exploring the potential effect of nonresponse on survey variables is to use reluctant respondents as proxies for final refusals. This approach is exemplified by Smith (1984) who divided

respondents into cooperative respondents (who willingly agree to the interview) and temporary refusals (who decline at one or more points but later agree to an interview). The assumption is that reluctant respondents are more similar to final refusals than they are to amenable respondents, and can thus be used as substitutes for those never interviewed. This approach has been widely used to evaluate nonresponse bias (see, for example, O'Neil 1979; Stinchcombe, Jones and Sheatsley 1981).

Smith (1984) found no noticeable differences between cooperatives and temporary refusals on a wide variety of attitudinal measures from the 1982 General Social Survey. If true, these findings have two implications: (1) there does not appear to be much bias from nonresponse, and (2) adjustments for nonresponse based on response propensity (i.e., reluctance) should not have much effect on estimates of substantive variables.

On the other hand, Stinchcombe et al. (1981) found "substantial differences" between temporary refusers and cooperative respondents in a survey of Dakota farm operators. In a telephone survey in the city of Chicago, O'Neil (1979) found that amenables and refusers differed, not only on demographic characteristics, but also on substantive and attitudinal variables (in this case issues related to crime). To what extent do Smith's findings represent an anomaly? Our data are similar to Smith's (1984) in that they are based on a national probability sample using face-to-face interviews, collecting measures of political attitudes and behaviour, permitting a re-evaluation of this issue.

In addition to being a partial replication of Smith's study, this paper also extends this work to evaluate the use of other variables to serve as indicators of possible nonresponse bias. Typically, most surveys make use of available frame (e.g., region, urbanicity) or process variables (temporary refusals, number of calls) as indicators of response propensity. We have been working to enrich this set of variables, both for modeling response propensity, and for use in nonresponse adjustment. This has involved the collection of additional observational measures during the data collection phase that may inform our understanding of the likelihood of participation and its effect on substantive variables (see Groves and Couper 1995a, 1995b).

2. Design and Data Collection

The data are from the 1990 National Election Study (NES) conducted by the Survey Research Center for the Center for Political Studies (Miller, Kinder, and Rosenstone 1992a). This is a nationally representative sample, using personal visits to attempt interviews from 2,806 eligible sample persons. At the end of each contact with a sample person, interviewers completed a series of items, providing information on what was said by the sample

person and interviewer, and other characteristics of the contact. These and other data from the administrative file (Miller, Kinder, and Rosenstone 1992b) were merged to the full dataset.

While this approach does not capture the full richness of these introductory interactions, it nonetheless provides summary information on a large number of contacts. For further details of this approach, and some evidence of the validity and reliability of these measures, see Groves and Couper (1995b). Interviewer actions and statements during the introductory conversation are also recorded on the contact description forms; however, we focus in this paper on respondent statements only.

The data from the contact description questions (one record per contact), and detailed call record data were aggregated and merged onto the full (case level) dataset. The final dispositions of the 2,806 eligible sample units in the 1990 NES was as follows:

	Number	**Percent**
Interview	2,000	71.3 %
Refusal	571	20.3 %
Noncontact	71	2.5 %
Other noninterview	164	5.8 %

For all but 3 of the interviewed cases, and for all but 13 of the refusal cases, contact description forms were completed by interviewers for at least one contact. Because of the small number of noncontacts and the unavailability of contact description information for this group, noncontacts and ":other noninterviews" (those unable to participate for reasons of language, incapacity, etc.) are excluded from these analyses. Thus, the analyses are limited to the 1,997 interviewed cases and 558 refusal cases for which contact description information was obtained.

The data for interviewed cases are weighted by the number of eligible adults in the household. Standard error estimates and statistical tests presented in this paper are calculated using Taylor Series approximation, reflecting stratification and clustering of the survey design (using SUDAAN, Shah et al. 1993).

On the basis of analyses performed earlier on the 1990 NES (see Couper 1995), we focus attention on two statements made by respondents during the introductory conversation (the request for participation). The most frequently made comments by sample persons in response to the request for participation relate to time constraints ("I'm too busy", "I don't have time") and lack of interest in the survey topic ("I'm not interested (in politics)", "I don't know anything about politics"). We refer to these sets of statements as "too busy" and "not interested" respectively.

As indicators for reluctance we use two other measures similar to those used by Smith and others. These are a refusal conversion indicator (for those cases which were coded as an initial refusal at any time) and a persuasion letter indicator (such letters being sent at the interviewer's request to reluctant respondents).

3. Analyses

The cooperation rates for these four indicators of reluctance are presented in Table 1. While all are statistically significant, the refusal conversion indicator and the persuasion letter are clearly powerful predictors of eventual cooperation. Note the relative infrequency of initial refusals (7.1 %). The next two indicators are components of the time-delay and negative statements used elsewhere (Groves and Couper 1995b; Groves, Raghunathan and Couper 1995). This table suggests that these measures (collected on both respondents and nonrespondents) are useful indicators of response propensity.

Table 1. Relationship Between Indicators of Reluctance and Survey Cooperation.

	Cooperation rate	(n)	Column Percent (% Yes)
All Cases	78.2%	(2,555)	
Refusal conversion indicator			
Yes	37.0%	(181)	7.1%
No	81.3%	(2,374)	
Persuasion letter sent			
Yes	40.8%	(789)	30.9%
No	94.9%	(1,766)	
"Too busy"			
Yes	61.9%	(766)	30.0%
No	85.1%	(1,789)	
"Not interested"			
Yes	71.8%	(510)	20.0%
No	79.8%	(2,045)	
"How long will it take?"			
Yes	89.0%	(824)	32.3%
No	73.0%	(1,731)	

We next address the question whether these variables are associated with any of the key substantive variables. Table 2 presents X^2 values for the associations for each of these four variables on a subset of key variables from the NES. The first two columns generally support the finding of Smith (1984) that there is not much relationship between reluctance and substantive responses. Only 4 of the 20 measures tested reach significance ($p<.05$) for refusal conversion and 2 for the persuasion letter indicator.

Table 2. Effect of Reluctance Indicators on Substantive Responses: X^2 Values.

	Refusal conversion	Persuasion letter	Mention lack of time	Mention lack of interest
Interest in the political campaign	3.62	0.19	1.18	44.3**
Attention to campaign in media:				
Read about campaign in newspaper	5.21*	0.52	1.12	21.0**
Watch campaign programs on TV	3.18	3.53	5.19*	10.1**
Discuss politics with friends	0.41	0.26	0.00	14.7**
Care about outcome of congressional election?	1.87	4.29	3.02	28.6**
Remember house candidates names?	0.32	9.94**	0.87	24.6**
Self-reported vote in 1990 election	0.42	0.67	0.37	24.5**
Political activity (other than voting)	0.07	0.05	0.00	7.68**
Follow government and public affairs	0.88	1.73	7.43	32.4**
Public officials don't care much what people like me think	0.77	0.18	7.29*	13.2**
People like me don't have a say in what the government does	0.65	0.85	3.14	13.3**

A similar pattern is found for those who mention lack of time in response to the interviewer's request for an interview. An initial examination of these three columns might lead to the conclusion that nonresponse bias is not of great concern in the NES.

However, the last column in Table 2 shows quite a different picture. Those respondents who say "not interested" during the initial request for an interview, are significantly different for all of the variables presented than those who do not mention a lack of interest in the topic/survey. The direction of these effects (not shown here) is consistent with expectation: those who express disinterest in the initial interaction have lower levels of political interest, participation, knowledge and more alienated attitudes. This finding is explored in more detail elsewhere (Couper 1995). We focus here on the implications of these results for nonresponse.

What these findings suggest is that the best predictors of response propensity (reluctance and persuasion letters) are not good indicators of differences in substantive variables. Thus we would expect the effects of nonresponse adjustment using these variables to be minimal. Similarly, a variable such as "not interested", while a strong indicator of substantive differences in responses to later questionnaire items, is a less powerful predictor of response propensity.

The next step is to use each of the variables in turn to form weighting classes for nonresponse adjustment, and evaluate their effect on key estimates. These are crude weights for illustrative purposes: any real weighting class scheme would cross-classify sets of variables to produce a larger number of weighting classes. For each of the variables we only produce two weighting classes (e.g., reluctant/not reluctant). We also note that the public use NES data file did not in 1990 (and still does not) include

Table 3. Effect of Various Adjustments on Estimates.

	Selection weights only	Alternative Nonresponse Adjustments: Refusal conversion weight	Persuasion letter weight	Too busy weight	Not interested weight	Combined regression weight
Interest in the political campaign (% very much interested):	20.6	20.6	20.6	20.4	20.4	20.4
Attention to campaign in media (% yes):						
Read about campaign in newspaper	65.6	65.9	65.8	65.7	65.2	66.0
Watch campaign programs on TV	62.9	62.6	63.8	62.5	62.7	63.2
Discuss politics with friends	69.4	69.6	69.6	69.4	69.0	69.1
To what extent do you care about the outcome of the congressional election (% very much)?	15.9	15.9	15.5	15.5	15.7	15.2
Remember house candidates names (% yes)?	30.7	30.6	29.2	30.5	30.4	28.6
Self-reported vote in 1990 election (% yes):	47.1	47.0	46.6	47.2	46.6	46.3
Political activity (% one or more activities):	55.7	55.6	55.8	55.7	55.5	55.3
Follow government and public affairs (% most of the time):	27.1	27.0	27.0	26.8	26.8	26.8
Public officials don't care much what people like me think (% agree)	64.0	64.1	64.1	64.1	64.2	64.1
People like me don't have a say in what the government does (% agree)	54.1	53.9	53.9	53.8	54.3	53.4

adjustments for nonresponse. The proposed nonresponse adjustment weights for the 1994 NES (Survey Research Center, 1995) are based on the formation of 19 weighting cells formed by cross-classifying region (9 census divisions) by PSU type (self-representing, nonself-representing MSA or nonMSA).

One of the problems of evaluating nonresponse adjustment weights in reducing bias is that the direction of the bias is usually unknown. However, in the case of the NES, there is evidence that the survey may over-estimate political participation, interest and knowledge. Hence, we can use the hypothesised direction of the effects to evaluate the adjustments. We assume that nonrespondents are likely to have lower levels of interest, participation, etc., and that the adjustments should produce reduced estimates of these variables.

Table 3 presents the effects of adjustments for each of these four variables in turn on the same set of estimates used earlier, with the unadjusted estimates in Column 1. The final column contains response

Table 4. Effect of Various Adjustments on Model of Electoral Participation.

Predictors	Selection weights only	Alternative Nonresponse Adjustments: Refusal conversion weight	Persuasion letter weight	Too busy weight	Not interested weight	Combined regression weight
Intercept	-3.44**	-3.49**	-2.91*	-3.49**	-3.45**	-2.65
Education	0.13*	0.13**	0.092	0.12*	0.13*	0.071
Age	0.037**	0.038**	0.036**	0.038**	0.037**	0.037**
Black	-0.26	-0.27	-0.25	-0.27	-0.25	-0.28
Hispanic	-0.13	-0.078	-0.23	-0.14	-0.12	-0.20
Income	0.0056*	0.0046	0.0039	0.0048*	0.0054*	0.0031

propensity weights obtained from a logistic regression model using these four variables, plus an additional one (the log of the number of calls to the sample household). First, looking at the bivariate weight-class adjustments (Columns 2 through 5), it can be seen that the adjustments do not have much effect on these estimates. In fact, the average absolute change is 0.17 % for the refusal conversion weight, 0.42 % for the persuasion letter weight, 0.15 % for the "too busy" weight, and 0.22 % for the "not interested" weight for the 20 variables evaluated. Thus, using each of these variables singly as adjustment factors produces little change in the estimates. Furthermore, for the first three adjustments, the direction of the change is in the hypothesised direction for only about half the cases. In contrast, 18 of the 20 estimates move in the expected direction for the "not interested" weight, although the overall effect is not large.

The adjusted estimates in the final column represents the combined effect of reluctance measures obtained from production data, and the two indicators of reactions to the topic from the contact observation measures. The average absolute change in estimates for this adjustment is 0.70 %, higher than each of the individual weights. Some of the effects are quite large: a reduction of 2.1 % in the proportion who remember the names of the house candidates running in their district, a reduction of 1.4 % in the proportion of validated voters (among all respondents), and a reduction of 0.8 % in the proportion of self-reported voters. However, the direction of the effects is not consistent (only 11 of 20 are in the expected direction).

Do the adjustments have greater impact on subgroup estimates, or on relationships among variables? We examined the same set of 20 variables for a number of key subgroups, including Blacks (n = 271), those 35 years old or younger (n = 758), those over 65 (n = 216) and those with less than high school education (n = 474). In general, the effects for the combined regression weight are somewhat larger: the absolute average differences across these 20 variables are 1.1 % for those over 65, 1.05 % for Blacks, 0.75 % for those 35 or younger, and 0.65 % for those with low education. Furthermore, some of the individual changes in estimates are quite large; for

example, a reduction of 3.1 % of Blacks and 3.4 % of elderly who recognize Margaret Thatcher, a reduction of 2.4 % of Blacks who watch campaign programs on TV, and an increase of 2.6 % of those with low education who agree with the statement "people like me don't have a say in what the government does." For Blacks the effects are in the expected direction for 16 of the 20 estimates, while it is less than that for the other groups examined.

We also examined the effects of these alternative adjustments on two models of self-reported voting behaviour commonly found in the literature. The first model (shown in Table 4) includes only demographic variables as predictors, while the second (not shown) adds attitudinal measures such as partisanship, level of interest in politics, and perceived complexity of the political situation. Similar results are obtained for the two models. It can be seen that the combined regression weight leads to different conclusions about the effects of certain covariates if traditional levels of statistical significance ($p<.05$) are used. Both education and income are significant predictors of self-reported voting behaviour in the unadjusted model; however, they fail to reach traditional levels of statistical significance ($p>.05$) under the combined regression weight. This drop in significance is primarily due to the reduction in the size of the coefficients rather than an increase in the variance due to the weights.

While the overall effects of the alternative adjustments on univariate estimates for the full population are neither large nor generally consistent, the adjustments do affect certain key variables in hypothesised directions. Furthermore, the effects of adjustment appear to be somewhat larger and more consistent when subgroups are analysed.

4. Conclusions

We have shown that those who say "not interested" prior to agreeing to participate in the interview differ on a number of key substantive indicators from those who don't make such statements. However, we still don't know whether or how those who say "not interested" and never participate differ from those who do make these statements but eventually cooperate. Without external data, this cannot be known. Still, these findings suggest that it might not be sufficient to claim a lack of nonresponse bias by showing that reluctants and amenables share similar characteristics.

We have also shown that measuring correlates of response propensity is not sufficient. It is important to also measure correlates of the substantive measures of interest. This may be especially true when potential respondents attempt to disqualify themselves from participation on the basis of the survey topic (e.g., lack of knowledge or interest for political surveys, lack of strong opinions on attitude surveys, age or ill-health for surveys of the elderly, good

(or bad) health for health surveys, etc.). Thus, our measures should not only include observable socio-demographics of the household (that may or may not be related to propensity or key substantive variables) and indicators of reluctance or effort (that are related to the former, but probably not the latter), but also information provided by the sample person during the initial interactions.

While the effect of the adjustments are generally modest for the few variables and subgroups we have examined, the crude weights we use here nonetheless affect key estimates in expected directions. This suggests that nonresponse bias is of concern, and that nonresponse adjustments based on measures of propensity and indicators of substantive variables are useful in reducing the impact of such bias. Obviously, much work can be done in the improvement of the response propensity models, using approaches similar to that discussed by Groves, Raghunathan and Couper (1995). Nonetheless, these analyses have shown that this is a fruitful line of research.

References

Couper, M. P. (1995). The Impact of Survey Introductions on Data Quality. *Paper presented at the International Conference on Survey Measurement and Process Quality, Bristol, April.*

Elliot, D. (1991). *Weighting for Non-Response; A Survey Researcher's Guide.* London: Office of Population Censuses and Surveys, Social Survey Division.

Groves, R. M. and Couper, M. P. (1995a). Theoretical Motivation for Post-Survey Nonresponse Adjustment in Household Surveys. *Journal of Official Statistics*, 11 (1): 93–106.

Groves, R. M. and Couper, M. P. (1995b). Contact-Level Influences on Cooperation in Face-to-Face Surveys. *Journal of Official Statistics, forthcoming.*

Groves, R. M., Raghunathan, T. E. and Couper, M. P. (1995). Evaluating Statistical Adjustments for Unit Nonresponse in a Survey of the Elderly. *Paper presented at the Sixth International Workshop on Household Survey Nonresponse, Helsinki, October.*

Miller, W. E., Kinder, D. R., Rosenstone, S. J. and the National Election Studies (1992a). *American National Election Study, 1990: Post-Election Survey.* Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

Miller, W. E., Kinder, D. R., Rosenstone, S. J. and the National Election Studies (1992b). American National Election Study, 1990: *Post-Election Survey (Field Administrative File).* Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

O'Neil, M. J. (1979). Estimating the Nonresponse Bias Due to Refusals in Telephone Surveys. *Public Opinion Quarterly*, 43: 218–232.

Smith, T. W. (1984). Estimating Nonresponse Bias with Temporary Refusals. *Sociological Perspectives*, 27 (4): 473–489.

Stinchcombe, A. L., Jones, C. and Sheatsley, P. (1981). Nonresponse Bias for Attitude Questions. *Public Opinion Quarterly*, 45: 359–375.

Survey Research Center (1995). Technical Description; 1990 National Election Study Sample Design. Ann Arbor: Survey Research Center, Sampling Section (Internal Memorandum).

THE EFFECT OF THE INTERVIEWER ON THE DECISION TO COOPERATE IN A SURVEY OF THE ELDERLY

Edith de Leeuw and Joop Hox
Department of Social Research Methodology,
Vrije Universiteit
Department of Education, University of Amsterdam

Survey nonresponse is a growing problem in Western Europe and the US. In a survey interview two actors are involved: the respondent and the interviewer. Past research on survey nonresponse has shown that one of the few consistent respondent characteristics associated with nonresponse is age. Furthermore, large differences in obtained response rates have been observed between interviewers. In 1992 a large survey of the elderly (55–89) was performed in the Netherlands. This survey provided us with biographical data on both respondents and nonrespondents. Also the degree of cooperation was registered of every contacted sample member (i.e., eager to cooperate, reluctant but cooperated after some discussion, refused after some discussion, refused immediately). In addition data on the interviewers were collected. Both biographical data, job-related attributes, and personality characteristics and social skills were available. A multilevel analysis with two levels (respondents and interviewers) was performed to explain degree of survey cooperation.

Key words: Nonresponse, degree of cooperation, decision process, respondent characteristics, interviewer effect, elderly, multilevel research.

1. Introduction

The growing concern about nonparticipation in surveys and polls (cf. Lyberg 1991; Bradburn 1992; Smith 1994), has stimulated research into the causes of nonresponse. Studies about the characteristics of nonrespondents repeatedly showed that age is an important indicator of nonresponse. Age as an individual variable is showing a consistent pattern with a lower response among older people. For a review see e.g., Goyder (1987), Groves (1989), Herzog and Rodgers (1988). When surveying elderly in the Netherlands, we explicitly incorporated nonresponse analysis in the NESTOR-LSN research program.

In interview surveys the interviewer plays an important role in gaining cooperation. There is empirical evidence that there is a considerable variation in response rates between interviewers (Lyberg and Lyberg 1991; Lyberg and Dean 1992). Despite this, research on the role of the interviewer in nonresponse is sparse. Several studies have addressed interviewer experience, but the results are mixed; for an overview see Couper and Groves (1992). In this study we systematically investigate the role of interviewer characteristics in unit nonresponse. Recent studies (e.g., Morton-Williams 1994) emphasize that the decision to cooperate in a survey is the outcome of a dynamic decision process: some respondents can by coerced by successful interviewers to participate after a hesitation or even an initial refusal. Therefore, we concentrated on the *degree of cooperation*, the twilight zone between response and nonresponse. The special structure of the data in interviewer research – respondents are nested within interviewers – suggests that a multilevel model should be applied to make statistically correct inferences (cf. Hox 1994, 1995).

In this paper we will address the following two questions using multilevel models: which respondent can be influenced and which type of interviewer does best.

2. Method

The data used in this study were collected in study on living arrangements and social networks of older adults (cf. Broese van Groenou, Van Tilburg, De Leeuw and Liefbroer 1995). Data on prospective respondents and on interviewers were available from several sources.

For all sample units (respondents and nonrespondents) age and gender were obtained from the municipalities. The interviewers provided data on the final result of the approach (e.g., interview, refusal, noncontacted), and on the decision process (e.g., eager to cooperate, refused after some discussion). They also provided data on the housing (i.e., independent on one's own, independent but in housing units designated for elderly, house with special adaptations and provisions, nursing home or home for elderly). Based on the zipcode[1], data were provided on financial status, ownership house, value of property, and rent. These were combined into one variable – economic status – using homogeneity analyses (Homals, Gifi 1990). The variable economic status was categorised in five categories or pentiles.

[1] This information was made avaible by the Dutch company Geo-marktprofiel. It consists of aggregated information on Dutch zipcodes, which have on average a density of 15 households per zipcode.

On the interviewers' data were available from a self-administered questionnaire completed by the interviewers at the beginning of the training. These data consisted of sociodemographic data (i.e., age, gender, education, previous interview experience, previous interview training), scores on personality tests (i.e., extroversion, friendly disposition, conscientiousness, social anxiety, ability to efficiently terminate undesirable situations), and attitude toward elderly measured with a semantic differential. The field supervisors provided a global interviewer evaluation on the scale 1–10 (1 = very negative, 10 = very positive), based on audio taped interviews.

The interaction between interviewer and respondent is central in this study, so we excluded all noncontacts and concentrated on degree of cooperation after contact. Interviewers had noted the respondent's decision to cooperate on a 4-point scale: (1) refused immediately; (2) refused after discussion; (3) cooperated after discussion; and (4) cooperated immediately. Based on this scale five dichotomous variables were constructed which are indicators of decisions in survey cooperation. The variables are: 'refuser' contrasting the immediate refusers against all others (1 vs 234); 'responder' contrasting the immediate responders against the rest (4 vs 123); 'discusser' contrasting those who can be tempted into conversation against the immediate refusers (23 vs 1); 'converter' those who cooperate after discussion contrasted to those who do not (3 vs 2); and 'final response' the end result of the interaction (3 + 4 vs 1 + 2).

These variables were used as dependent variables in a multilevel analyses with the program VARCL. A logit link function was used to accommodate the binary dependent variables (cf. Longford 1990). For a detailed description of the statistical model and method used see Hox (1995, chap. 4.2).

The multilevel model had two levels: interviewer (n = 80) & respondent (N = 6151). The interviewer and respondent data described above were used as predictors. Dummy codes were used for age to indicate life cycle with as reference category the young elderly aged 55–59 (LC1 early retirement: 60–64; LC2 elderly: 70–84; LC3 very elderly: 85–89). Dummy codes were also used to indicate type of housing (H1: designated elderly; H2: special adaptations; H3: institution), the reference category used here was 'normal' independent housing.

The interviewer effects were assessed after respondent effects had been included. To control for imperfect randomisation of interviewers the geographical region was always included as control variable.

3. Results

The results of the analyses have been summarised in Table 1 where the final models for all five dependent variable – the indicators of the response process – are presented. Studying the significant respondent variables for each response indicator answers our first research question: which respondents refuse and which can be influenced. Studying the significant interviewer variables answers the second question: which interviewers do best. Special attention is given to cross-level interactions, the occurrence of these is an indication that certain types of interviewers do better with certain types of respondents.

Table 1. Final explanatory model for nonresponse process: dependent variables are refuse, respond, discuss, convert, and final response. Only significant explanatory variables are reported.

	Refuse	Respond	Discuss	Convert	Final
Fixed Eff.					
intercept	-1.32	0.49	-1.07	-0.35	0.68
region	-0.06	0.07	0.03	0.13	0.05
gender		-0.18			
H1	-0.25			0.37	0.30
H2	-0.65		0.65		0.21
H3	0.00[1]	-0.40		0.00[1]	-0.42
LC1				0.58	0.20
LC2					
LC3		-0.29			-0.25
LC4	0.29	-0.51			-0.57
SES	-0.12	0.11	0.06		0.09
Itr. vars					
education				0.15	
Formal/conscientious		0.22			
good eval		-0.23	0.23		
Rand. Eff.					
σ^2	1.00	1.00	1.00	1.00	1.00
σ^2interc null model	.19	.41	.32	.38	.09
σ^2interc final model	.12	.36	.28	.34	.04
σ^2H2					.12
σ^2H3	.58			.44	.24

[1] This value was not significant. Since the associated variance component is significant, the variable is kept in the model with its regression coefficient constrained to 0.00

To see which respondents immediately say NO after contact, we have to inspect the column 'refuse'. We note that the intercept variance of the basis or null model is 0.19, this corresponds to an intraclass correlation or interviewer effect of 0.16. (cf. Hox 1994). When respondent and interviewer variables are added, the intercept variance drops from 0.19 to 0.12. Those respondents who are very elderly (85–89), who are living independently, and have a low economic status are more prone to say no immediately at contact.

There are differences between interviewers in willingness for elderly living in institutions (random slope H3). None of the twelve interviewer variables available could explain this difference.

Column two of the table tells us who says immediately YES (responds) to an interviewer after contact. Male elderly, younger than 70, not living in institutions, and with a higher economic status are more inclined to respond immediately. They do this especially to interviewers who score high on conscientiousness (e.g., are organised, like to be on time, be properly dressed) and who are found to deviate sometimes from standard interviewer training. There were no random slopes, which means that there were no statistical cross level interactions. It should be noted that there is a large interviewer effect for this variable: the intercept variance for the null model is 0.41 (intraclass correlation = .29). The final model with respondent and interviewer variables together explains about 10 % of the initial variance.

Morton-Williams (1994) emphasizes the importance of doorstep interaction and discussion in converting initial refusers. Who can be tempted to start discussing cooperation with an interviewer at the doorstep? Column 3 labelled 'discuss' tells us that those elderly who live in designated adapted apartments and have a higher economic status are more easily tempted to enter into a discussion, especially when confronted with interviewers who were evaluated as good interviewers. Again there are large interviewer effects: the intercept variance for the null model is 0.32 (intraclass correlation = .24). The final model with respondent and interviewer variables together explains about 10 % of the initial variance.

After being engaged in a conversation with the interviewer, who will respond and who will refuse? Column 4 (convert) tells us that those in the age group 60–65 (early retirement) and living in designated apartments are more inclined to consent to an interview, especially with highly educated interviewers. There is a random slope for the variable designated apartment, indicating that certain interviewers are far better to convert elderly living in designated apartments than other interviewers. The intraclass correlation is 0.28 (intercept variance null model .38)

Although there are large interviewer differences in the different types of initial response – some are better in getting an immediate yes, other are better in engaging elderly in discussion and convert them into respondents – when we look at the final result the interviewer effects are relatively small. The intercept variance of the null model for final response is .09, indicating an intraclass correlation for interviewers of 0.08. Elderly younger than 65 and those not over 70, who live in designated or adapted apartments, and are not living in institutions, and with a higher economic status are more willing to respond. There are random slopes for special housing and institutions, indicating that some interviewers are better than others to get a favorable response with elderly who are living in special housing or are

institutionalised. Unfortunately non of the twelve interviewer variables available could explain the difference in final cooperation rates between interviewers.

4. Summary and Discussion

Although there are differences between interviewers their influence on the END-product is only limited: the interviewer variance of final cooperation was relatively low. The analyses confirm earlier research on nonrespondent characteristics. In general, refusers are old and have a low SES. There is a much larger initial interviewer variance; some interviewers can get an initial yes very easily, other interviewers are better in engaging the elderly in conversation and converting them. But concerning final response these effects almost equal out and the interviewers do only differ in a limited way. The intra class correlation is .08, which means that in the null model the proportion of interviewer variance is 8 %. However, in the final model, where known respondent characteristics are taken into account, the proportion of residual interviewer variance is only 4 %.

It is disappointing that the available interviewer variables (biographical data, personality scores, job related data, evaluation by supervisors) had hardly any significant contribution to the differences between interviewers. This is a confirmation of a smaller study by Hox, de Leeuw and Kreft (1992) with the general population.

Some observations however can be made. For instance, interviewers who scored high on the conscientious/formality items were better in getting an immediate yes. This means that interviewers who value being on time for appointments, who think it is important to dress correctly for the occasion, who adhere to norms, have perseverance and describe themselves as conscientious are better in getting a quick yes on the doorstep. Whether these interviewer attributes especially appeal to elderly respondents who are perhaps easier influenced by a reliable image, or are also important for the general populations is not sure and should be investigated. A second striking result is that higher educated interviewers were far better in engaging hesitant respondents into a doorstep conversation about the study. Morton-Williams (1994) shows that doorstep conversations are a positive sign and often result in a response. In this light the above results can be useful both in initial selection of higher educated and well-behaved interviewers and in training.

Finally, we have to comment on the apparent contradictory results concerning interviewers who are successful in getting a yes right away. Important interviewer variables were conscientiousness/formality and (negative!) evaluation. Those interviewers who were conscientious and behaved well and also were found by supervisors to deviate from the rules at

certain points, were best at the doorstep. Recent insights in coercion strategies emphasize the tailoring of behaviour to the situation at hand. When an interviewer is conscientious and wants to be polite and address persons in a proper way, he or she will be more likely to adapt quickly to the situation even if this means that they have to deviate from a prescribed script or from interviewer rules.

Acknowledgements: The data were collected in the context of the research program 'Living arrangements and social networks of older adults.' This research program is conducted at the departments of Sociology & Social Gerontology and Social Research Methodology of the Vrije Universiteit in Amsterdam, and the Netherlands Interdisciplinary Demographic Institute in the Hague and is supported by a program grant from the Netherlands Program for Research on Aging (NESTOR), funded by the Ministry of Education and Sciences and the Ministry of Welfare, Health and Cultural Affairs.

References

Bradburn, N. M. (1992). A response to the nonresponse problem. *Public Opinion Quarterly*, 56, 391–397.

Broese van Groenou, M., Van Tilburg, T., De Leeuw, E. and Liefbroer, A. (1995). Data collection. In: C. P. M. Knipscheer, J. De Jong Gierveld, T. G. Van Tilburg, & P. A. Dykstra, (Eds). *Living arrangements and social networks of older adults*. Amsterdam: VU University Press.

Couper, M. P. and Groves, R. M. (1992). The role of the interviewer in survey participation. *Survey Methodology*, 18, 263–277.

Gifi, A. (1990). *Nonlinear multivariate analysis*. New York: Wiley.

Goyder, J. (1987). *The silent minority; Nonrespondents on sample surveys*. Cambridge: Policy Press.

Groves, R. M. (1989). *Survey error and survey costs*. New York: JohnWiley

Herzog, A. R. and Rodgers, W. L. (1988). Age and response rates to interview sample surveys. *Journal of Gerontology: Social Sciences*, 43, 200–205.

Hox, J. J. (1994). Hierarchical regression models for interviewer and respondent effects. *Sociological Methods and Research*, 22, 300–318.

Hox, J. J. (1995). *Applied multilevel analysis*, 2nd edition, Amsterdam: TT-publikaties.

Lyberg, I. and Lyberg, L. (1991). Nonresponse research at Statistics Sweden. *Invited paper presented at the Annual Meeting of the American Statistical Association, Atlanta.*

Lyberg, I. and Dean, P. (1992). Methods for reducing nonresponse rates: A review. *Paper presented at the Annual Meeting of the American Association for Public Opinion Research. St. Petersburg, Fl.*

Morton-Williams, J. (1994). A social skills analysis of interviewer doorstep strategies. *Paper presented at the 4th International Workshop on Household Survey Nonresponse, Bath, UK.*

Smith, T. W. (1994). Changes in non-response on the US General Social Survey, 1975–1994. *Paper presented at the 5th International Workshop on Household Survey Nonresponse, Ottawa, Canada.*

THE FIRST YEAR OF COMPUTER-ASSISTED INTERVIEWING FOR THE CANADIAN LABOUR FORCE SURVEY: AN UPDATE

Johane Dufour, Michelle Simard and Frank Mayda
Statistics Canada, Canada

As a part of its decennial redesign, the Canadian Labour Force Survey (LFS) has recently introduced new technology in data collection. Computer-Assisted Interviewing (CAI) was introduced over a period of four months starting in November 1993. Since March 1994, CAI has replaced traditional paper and pencil data collection. This paper describes the impact of this major change on traditional quality indicators, such as nonresponse, that have been observed in the first year. The paper also introduces some new data quality indicators for the interview process made possible by CAI. Finally, the paper points to lessons learned from the change in process and identifies future directions.

Key words: Computer-Assisted Interviewing, data collection, data quality indicator, nonresponse.

1. Introduction

The Canadian Labour Force Survey (LFS) has recently undergone a major technological changeover. Computer-Assisted Interviewing (CAI) is now the data collection method. Beginning in November 1993, the CAI mode was gradually introduced in the LFS. Within a few months, portable computers (notebooks) replaced the traditional Paper And Pencil Interviewing (PAPI). An overview of how such a fundamental change to the survey's methodology was achieved is given in Dufour, Kaushal, Clark and Bench (1995). This paper focuses on the impact of this major change on some data quality indicators, mainly on the nonresponse rate, as well as the challenges encountered during the implementation. It also discusses the introduction of new types of quality indicators which are now available with the implementation of CAI.

This paper is divided into six sections. The following section briefly describes the Canadian Labour Force Survey. The third section discusses the conversion strategy from PAPI to CAI. The fourth section analyses current

and new data quality indicators. The current data quality indicators are measures which are regularly produced such as nonresponse rates and vacancy rates. New data quality indicators can be divided into two types. There are those which were produced since the introduction of CAI to monitor and measure the performance of CAI. As well, there are the quality indicators from the case management system of CAI which provide previously unavailable information about the interview process, such as the average duration of the interview and the number of calls required to contact the respondent. The last two sections outline the lessons learned and the future of CAI for the LFS.

2. The Canadian Labour Force Survey: An Overview of the Survey

The LFS is the largest ongoing household survey conducted by Statistics Canada. This survey produces monthly estimates of labour force characteristics of the Canadian population, such as employment and unemployment at national and provincial levels, as well as by industry and occupation. The LFS uses a stratified multi-stage sampling plan with the dwelling as the final sampling unit. The sample is split into six representative sub-samples or panels, and each month the dwellings from one of the panels (one-sixth of the sample) are replaced. Selected dwellings remain in the survey for six consecutive months. Approximately 55,000 households, representing about 110,000 individuals, are in the sample each month (see Singh et al. 1990).

Statistics Canada's six Regional Offices (RO) employ 950 interviewers to conduct LFS interviews. The first (or birth) interview with the household in the dwelling is conducted in person by an interviewer. Subsequent interviews are conducted by telephone. Prior to November 1993, all interview results were recorded using paper and pencil, entered into minicomputers at the RO and subsequently edited at the Head Office.

3. Conversion Strategy: from PAPI to CAI

Even though the implementation of CAI implied complex reorganization and fundamental restructuring of survey processes, there were advantages to converting: (i) improvement in data quality (through on-line editing done directly with the respondent and elimination of human errors such as not following skip and branching patterns in questionnaires); (ii) faster data processing (data capture is now done by the interviewer at the time of the interview); (iii) long-term cost benefits (due to combining the interviewing

and data capture steps); (iv) development of a generalized data collection tool and (v) the possibility of handling more complex questionnaires (internal programming of all questionnaire skip and branching patterns into the notebook lets it automatically display the next relevant question). One of the first major impacts of CAI will take place in 1997, when a new questionnaire will be introduced. This questionnaire will be much more complex than would have been possible with PAPI.

Before actually implementing CAI, there were several years of evaluation and testing. During the late 1980's, the LFS has undergone three major tests, with the specific aim of evaluating the potential of computerized data collection. The principal objectives of the first two tests were: to test the potential of CAI (Catlin and Ingram 1988) and to test the feasibility of using new technologies in the LFS (Kaushal and Laniel 1993). The third test, called the Data Quality Test (Kaushal and Laniel 1995), was primarily aimed at evaluating the impact of the change on the continuity of LFS series' and on data quality. Operational development and evaluation of CAI were also covered in this test. The results of the third test concluded that CAI would have no discernible impact on either the continuity of data series' from the LFS or on the main quality indicators of the survey. On the basis of these findings, it was decided to convert the mode of data collection to CAI in the fall of 1993.

The strategy adopted was to gradually introduce CAI as the new collection method. This was done with three principal goals: (i) to avoid disruption in the historical series, (ii) to minimize the introduction of any collection mode bias and (iii) to minimize the massive change to the interviewer work procedures and in the data collection process. The strategy was to convert at random one third of the interviewers from PAPI to CAI, in all ROs in November 1993, another third in December 1993, maintain the status quo in January and February 1994 and finally to convert the remaining third in March 1994. There were two reasons for the status quo period: (i) to allow an adaptation period for adjustment and problem-solving and (ii) operational constraints. In January and February 1994, two new longitudinal surveys, which also required the use of the laptops, were introduced.

As part of the conversion strategy, when an interviewer was converted to CAI, his or her whole assignment had to be completed with CAI and that interviewer continued with the computer-assisted method during the succeeding months of the implementation process. All of the data collected with CAI were used in the production of LFS estimates. Paper and pencil interviewing was available as a contingency plan for all interviews which could not be completed by CAI for operational reasons. The PAPI backup option is in place until January 1997 when the new questionnaire becomes effective.

4. The impact of CAI on Data Quality Indicators

In this section, both current and new data quality indicators (DQI) are analyzed. The current DQI include some of the regular indicators which are produced and monitored monthly by the LFS Data Quality Committee (DQC), namely, nonresponse rates and vacancy rates. There are two types of new DQI: those which measure the performance of CAI (the conversion rate and the technical problems or "Z" codes) and those which are now available from the case management system. This system manages all of the survey activities from the beginning to the end of the survey cycle. Before analyzing these quality indicators, it is worth mentioning that since the introduction of CAI in November 1993, some changes have been made to the LFS that can have a direct impact on data quality indicators. For example, in September 1994, a new CAI application, i.e. a new software, was introduced. This application, faster than the previous one, implies a shorter duration of interviews and a reduction of respondent burden, which can have a direct impact on nonresponse. The second change was the introduction of the new sample. Since the new sample is more urbanized, it is expected that the nonresponse rate will be affected. Finally, between June and August 1995, a new version of the case management system was implemented. The main goal of this new version is to deal with problems in case transmission and consequently to reduce technical problems.

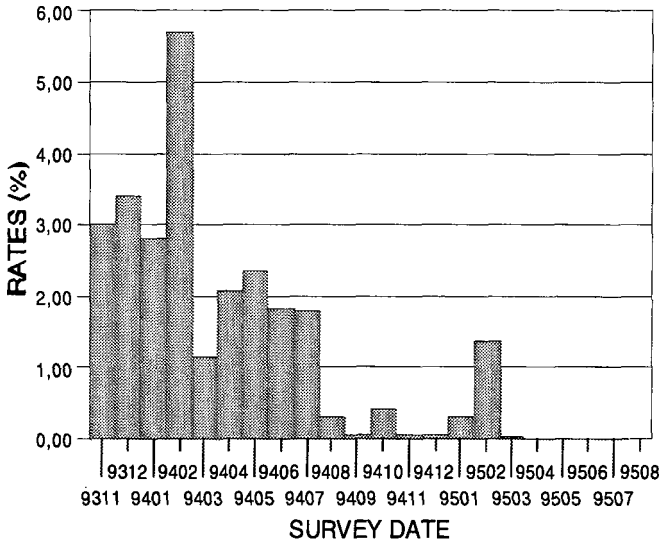
New Data Quality Indicators: Evaluation of the Performance of CAI

There are two indicators that are closely followed by the DQC to measure the effectiveness and efficiency of CAI. One of these is the conversion rate and the other is the technical problems or "Z" nonresponse code. Figure 1 shows the conversion rates from CAI to PAPI since November 1993. During the interview, interviewers who were not able to complete an interview with CAI, for any reason, could pursue the interview with PAPI. The proportion of interviews that should have been conducted with CAI but were converted to PAPI is reflected by the conversion rate. Ideally, there should be no conversion.

The conversion rate varied between 1.0 % and 5.7 % from November 1993 to July 1994 with an average of 2.6 % per month. From August 1994 until now, the conversion rate is always less than 0.5 %, except for the February 1995 survey (1.3 %). This high rate coincides with a PC communication failure which occurred in one Regional Office, and was responsible for 99 % of all cases converted in February. Since March 1995, the conversion rate has been zero.

Consequently, it took about a year for this rate to stabilize and reach a satisfactory level.

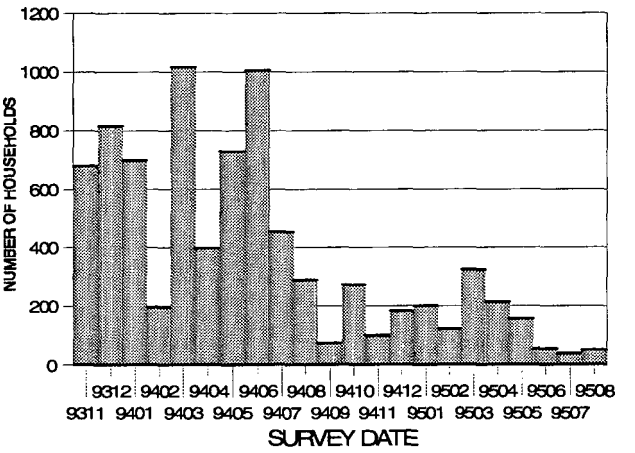
Fig. 1. Conversions in the field: from CAI to PAPI.



As for the second indicator, technical problems, the frequent occurrence of "Z"'s is a result of the use of new technology and was most unpredictable. With the introduction of CAI, "Z" codes became more frequent. Although this code existed previously, the definition has changed somewhat. "Z" nonresponse codes have been and are still defined as "too late for processing". Before CAI, they reflected a postal problem. Since the introduction of CAI, they are a reflection of technical problems such as: transmission problems, disruptions of telephone lines, uploading data processing system failures, automatic computer maintenance function which disconnects all transmissions without warning, etc. There were also hardware problems such as: hard disk failures, magnetic tape failures, insufficient memory allocations, etc. All these problems have been resolved case by case, requiring time and resources during the survey cycle. Since all LFS operations have a very tight schedule, sometimes, there was not enough time for these late records to be processed for the current survey.

As seen in Figure 2, the number of technical problems has decreased since the fall of 1994. This decrease coincides with a new version of the CAI application, which is largely responsible for this improvement. Like the conversion rate, it took about a year of CAI to observe an improvement in the "Z" codes. Prior to the introduction of CAI in November 1993, when "Z" codes represented failure to receive completed paper questionnaires on time, this rate was not worth monitoring closely. From November 1993 to August 1994, the average was about 600 cases per month, dropping to an average of 150 cases since September 1994. Moreover, for the last three survey months,

Fig. 2. Nonresponse: Number of technical problems.



the number of "Z"'s was less than 50 cases which coincides with the implementation of the new case management. The contribution in percentage of the "Z"'s to the nonresponse rate varied from 0.09 % to 1.74 %. It has stabilized around 0.1 % in recent surveys (see section 4.2).

These two indicators are used to assess the performance of CAI. Both measures showed relatively high rates and numbers at the beginning of the implementation of CAI. These two indicators show that it took about a year to stabilize.

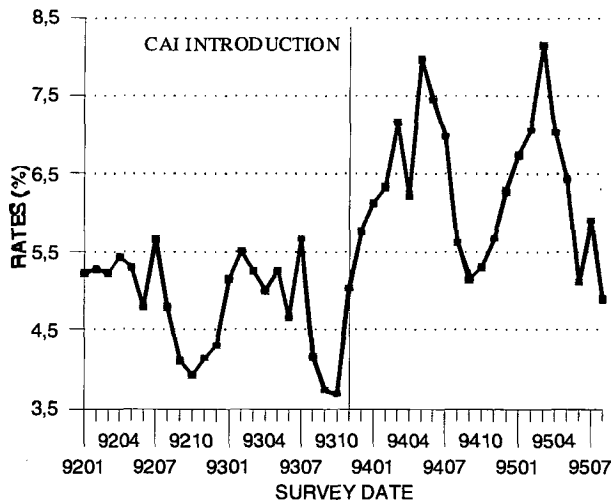
Regular Data Quality Indicators

Total Nonresponse rates During the four-month implementation period, from November 1993 to February 1994, both collection methods were used simultaneously. During these four months, nonresponse rates were systematically higher for the sample interviewed with CAI than with PAPI. (For a more complete look at this four-month implementation period see Simard and Dufour, 1995). There is a major difference between the nonresponse rate before and after the introduction of CAI. Figure 3 shows the total nonresponse rate for Canada since January 1992.

Historically, LFS nonresponse rates average around 5 % at the national level. During the first year of CAI, the nonresponse rate was generally higher than this. The national nonresponse series has a typical seasonal trend with peaks in summer months (usually in July), and troughs in the fall (most of the time in October). Since the introduction of CAI, this seasonal trend has been disturbed slightly. For example, two high values were recorded in May 1994 (7.8 %) and in March 1995 (8.1 %) rather than in the summer months, and the minimum for 1994 was observed in September rather than in October. Following the peak observed in March 1995, the

nonresponse rate has shown a downward trend. It took 21 months after the implementation of CAI for the nonresponse rate to reach its usual value.

Fig. 3. LFS nonresponse rates.



Another complicating factor is the new sample design which was introduced gradually over a six-month period starting in October 1994, as part of the major decennial redesign of the LFS. This new design has two features which affect nonresponse rates. One of the new characteristics is the greater proportion of urban to rural sample compared to the previous design. This fact does influence this analysis since it is well known that nonresponse is higher in urban areas than in rural areas (see Figure 4). The other feature is the related hiring of a number of new interviewers, who tend to have higher nonresponse rates than experienced interviewers (defined as 6 months or more of experience with the LFS), as observed in Figure 5. Therefore, part of the increase in nonresponse after October 1994 can be explained by the introduction of the new sample design.

As mentioned previously, another factor responsible for the increase in total nonresponse rates is the appearance of the technical problems; it is, in fact, the main reason. Figure 6 presents the effect of "Z" codes on nonresponse. After subtracting the portion of nonresponse due to "Z"'s from the total nonresponse, the curve shows patterns similar to the previous year.

During the four-month implementation period, the "Z" codes were the sole cause of higher nonresponse rate for CAI. After removing them, the average nonresponse rate was almost the same for CAI as PAPI (4.9 % vs. 4.7 %). From March 1994 to May 1995, the technical problems continually decreased. Since June 1995, the effect of the Z codes on nonresponse rate has been negligible; similar to what it was before the introduction of CAI.

Fig. 4. Nonresponse rates by area types.

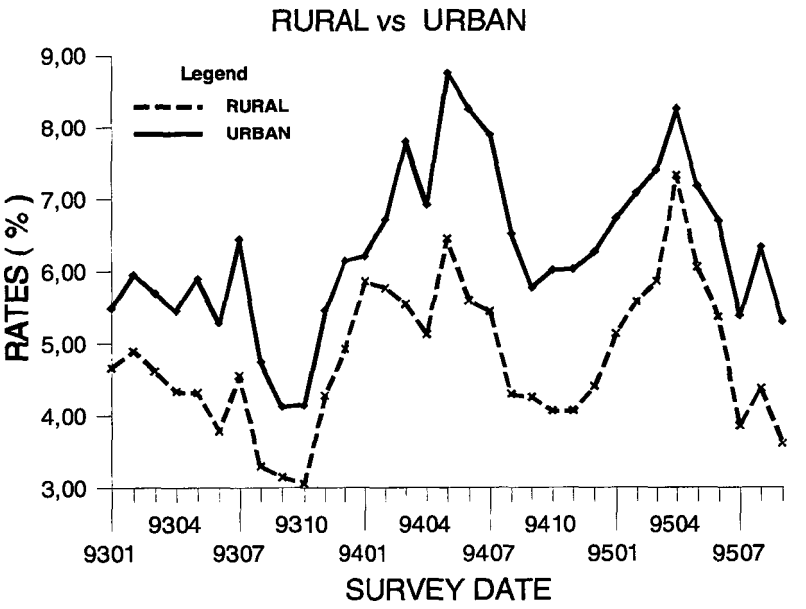


Fig. 5. LFS interviewer nonresponse rates by length of experience.

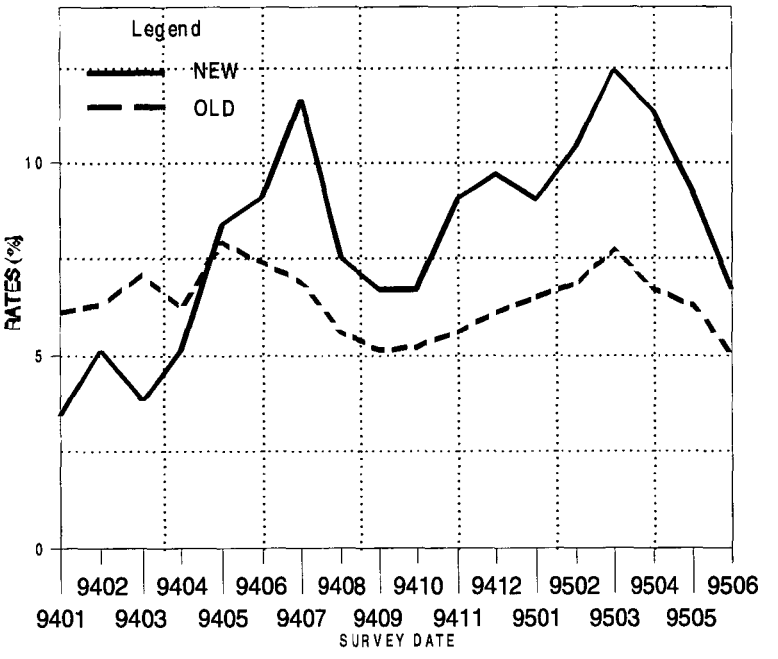
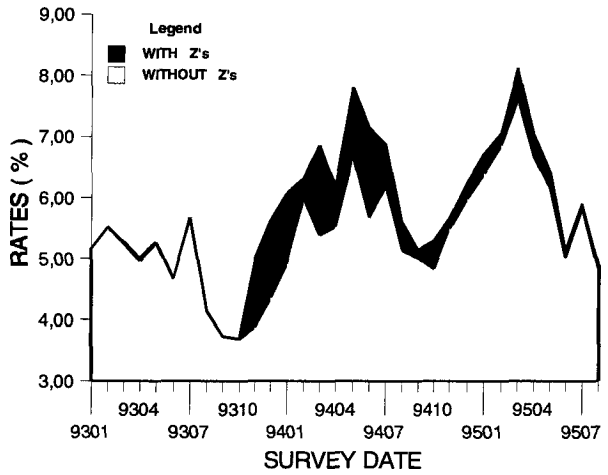


Fig. 6. Impact of technical problems on nonresponse rates.



Nonresponse by Reasons There are several reasons for LFS nonresponse, which are recorded by the interviewer: (i) household temporarily absent (T), (ii) no one at home (N), (iii) refusal (R), (iv) technical problems (Z) and (v) other reason such as no interview due to circumstances within the household. The latter category contributes very little to the nonresponse rate, rarely exceeding 1 %. Before the implementation of CAI, the total nonresponse rate was mostly a function of "Temporarily absent", "No one at home" and "Refusal". "Technical problems" had no role. However, with CAI, the total nonresponse rate is dominated by the incidence of "Technical problems", which became less numerous over time, and is less influenced by the other components. Figures 7, 8 and 9 present the nonresponse rates for the reasons T, N and R for January 1993 to August 1995.

Fig. 7. "Temporarily absent" nonresponse rate.

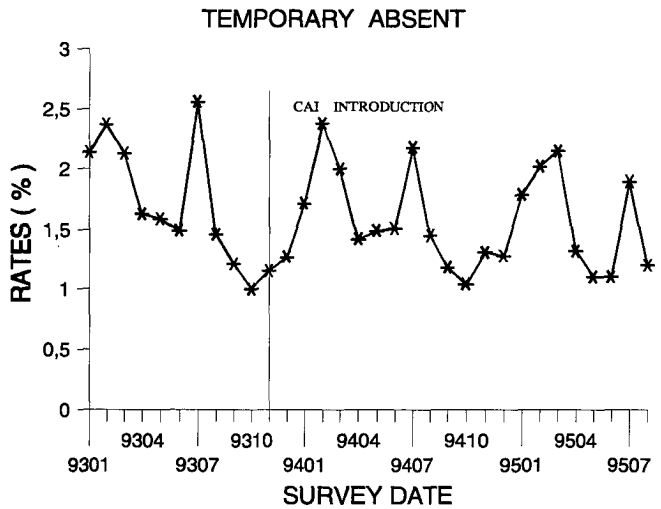


Fig. 8. "No one at home" nonresponse rate.

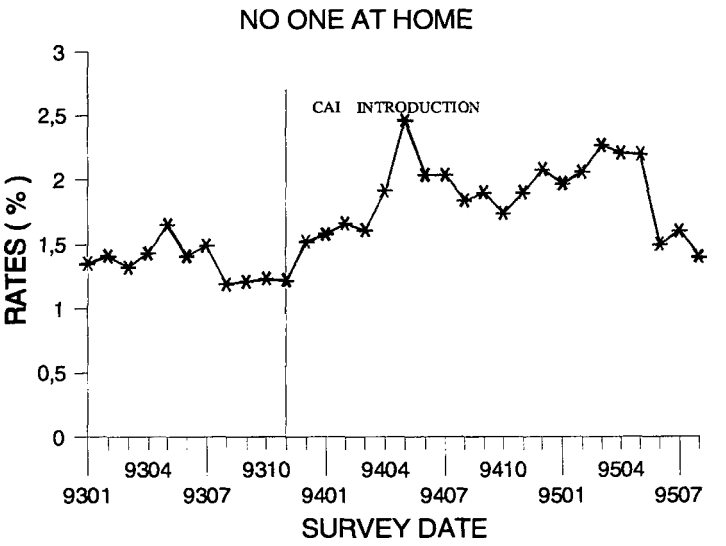
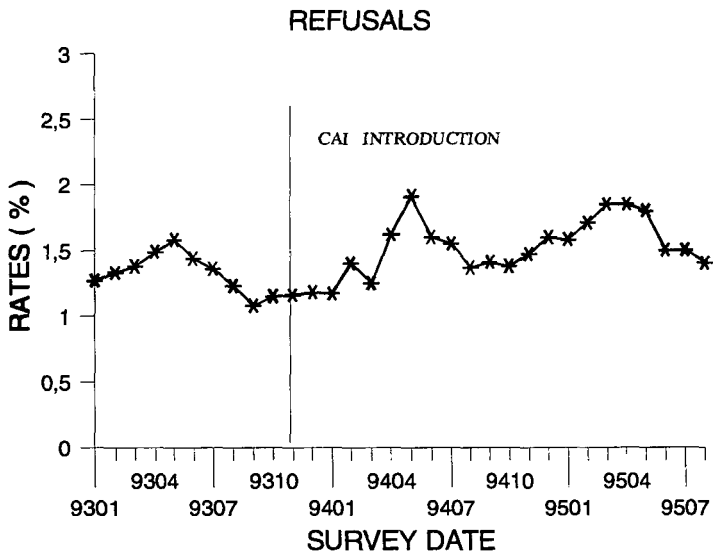


Fig. 9. "Refusal" nonresponse rates.

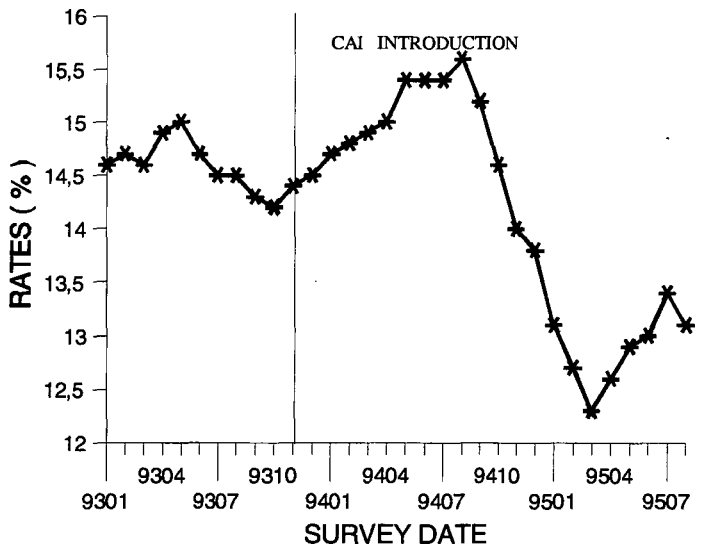


During the implementation period, these nonresponse reasons were almost the same under the two data collection modes. The introduction of CAI does not seem to have affected the behaviour of the temporarily absent series, nor does the new sample. As shown in Figure 7, high rates were obtained in February and July which coincides every year with winter school breaks and summer vacations. The "No one at home" reason (Figure 8) of nonresponse seems to be the most affected by the recent changes made to the LFS, mainly because the new sample design allocates more of the sample to urban areas.

The refusal rate (Figure 9) was quite stable, before and after the introduction of CAI, fluctuating between 1 % and 2 %. This component of nonresponse was one of the data quality measures for which LFS managers were concerned, given the presence of a computer in the household. Many felt that the respondent would be more reluctant to respond to the questionnaire than before. However, the nonresponse rates for this component with CAI do not show any major difference, when compared to the rates with PAPI.

Vacancy rates The vacancy rate for Canada is graphed in Figure 10. During the implementation of CAI, the vacancy rates for the CAI sample were systematically lower than the PAPI ones. The difference between the two modes decreased over the course of the implementation period. No specific cause for the difference has been identified. After the introduction of CAI, the national vacancy rate increased to reach 15.6 % in August 1994; its highest value since the previous redesign. Following this peak, a strong decrease was observed. This decrease is probably a consequence of the introduction of the new sample, which is more urbanized, and because of a more up to date sample frame.

Fig. 10. Vacancy rates.



New Data From the Case Management System

All computerized survey activities are managed by a sophisticated data management system called the Case Management System (CMS). The main functions of the CMS are: routing, reporting and providing assistance to the interviewers during the progress of the survey. From this system, new data can now be gathered and analyzed. Such information includes: the average

number of attempts before contacting the respondent, the best time and best day of the survey week to complete an interview and the average duration of interviews. A research team has been established to develop new data quality indicators and find efficient uses for them. These new indicators will allow a better understanding of the interviewers' work and will provide information for more efficient management of the survey. The new indicators can be used to improve interviewer training and reinforce good behaviour such as better assignment planning and scheduling. They will also provide quantitative information on many aspects of the interviewers' work.

Figure 11 graphs, for the August 1995 survey, the number of attempts before completing an interview by telephone and in person. The curves show that after 6 attempts to contact a household, the probability of reaching a respondent does not increase very much. The success rate (establishing contact) is about 95 % after 6 attempts for both types of interviews. After 10 contacts, the curves are stationary. By combining these data with cost data, an optimal number of attempts may be determined.

Fig. 11. Number of attempts to contact a household: Personal and telephone interviews.

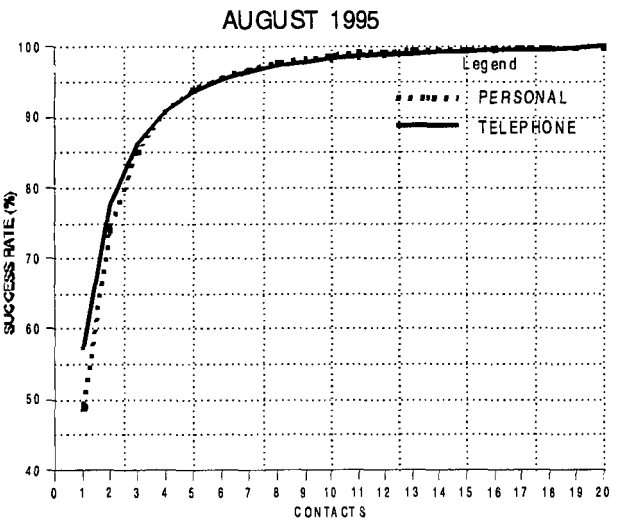
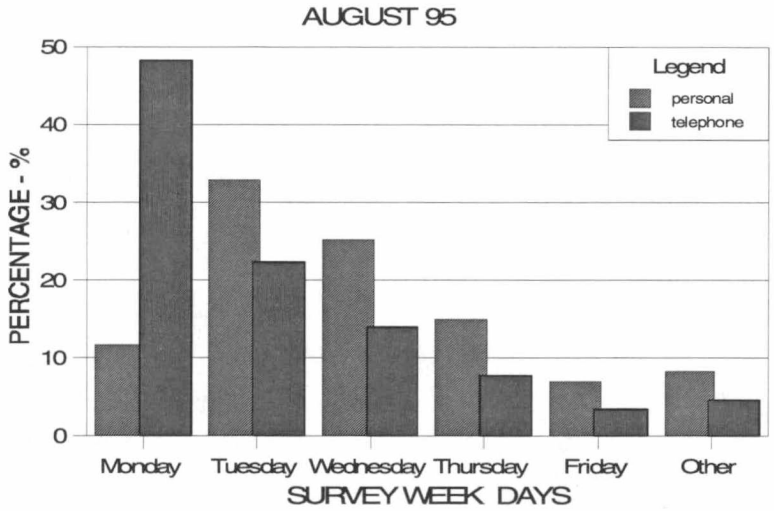


Figure 12 presents type of interview (personal and telephone) by collection day for the August 1995 LFS survey. About 70 % of telephone interviews are done during the first two days of the collection week. More specifically, the most frequent times on Monday to conduct a telephone interview are: 10 am to noon, 6 pm to 8 pm and 2 pm to 4 pm. The interviewers do most of their telephone interviews on Monday, and most of their personal interviews on all other days. Tuesday 2 pm to 4 pm, Wednesday 6 pm to 8 pm and

Tuesday 6 pm to 8 pm are the most frequent times for personal interviewing. The number of personal interviews decreases from Tuesday to Friday, but on Saturday, and during the summer months on the following Monday, additional efforts are made to reach respondents. During the summer, the collection week is increased by one day, and sometimes by 2 days since it is more difficult to contact respondents, especially those who are Temporarily Absent. This procedure was adopted to improve response rates.

Fig. 12. Distribution by collection day: Personal and telephone interviews.



The distribution of the day of the week on which a final code is allocated to the household is not uniform by nonresponse component. Some differences are expected since the treatment of different types of nonresponse is not the same. For example, interviewers are instructed to return as often as they can to a "No one at home" household, before finalizing the case. On the other hand, when nonresponse is caused by an unusual circumstance in the household, such as health problems or a death, no further contact is usually made. On average, an interviewer tries to contact a household by telephone 10 times for "Temporarily Absent", 11 times for "No One at Home" and 3 times for "Other" reasons. For personal interviews, the corresponding average numbers of contacts are: 7, 7 and 4. Figures 13 to 15 show these differences.

For "Temporarily Absent" households (Figure 13), most cases are finalized at the end of the collection period. Around 20 % of cases are closed between Thursday and Saturday (possibly confirmed by a neighbour, or the respondents may have told the interviewer in advance). The pattern is almost the same for personal and telephone interviews.

Fig. 13. Day of assigning the final code: "No one at home" nonresponse.

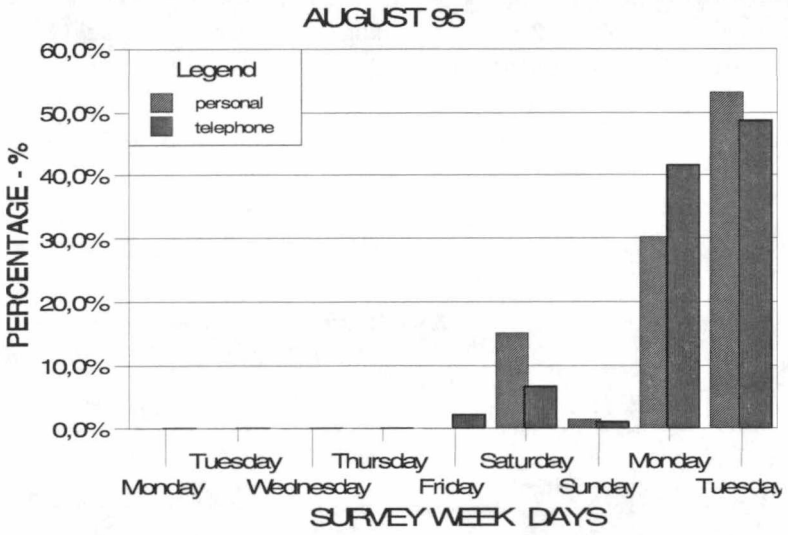
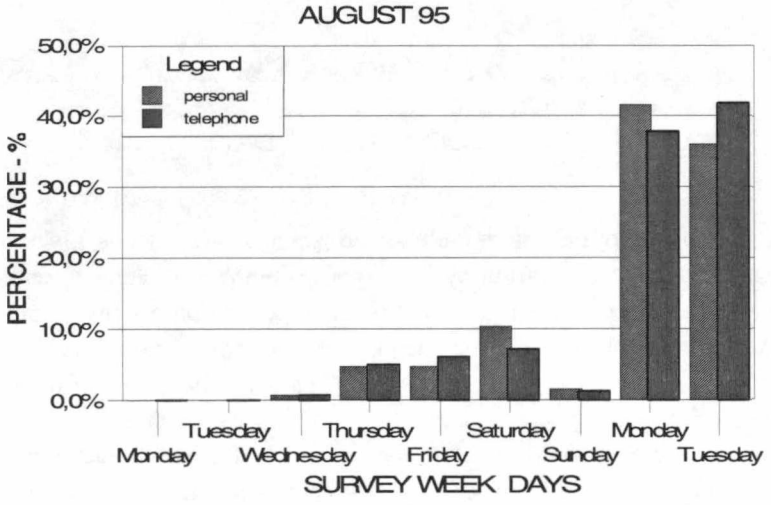


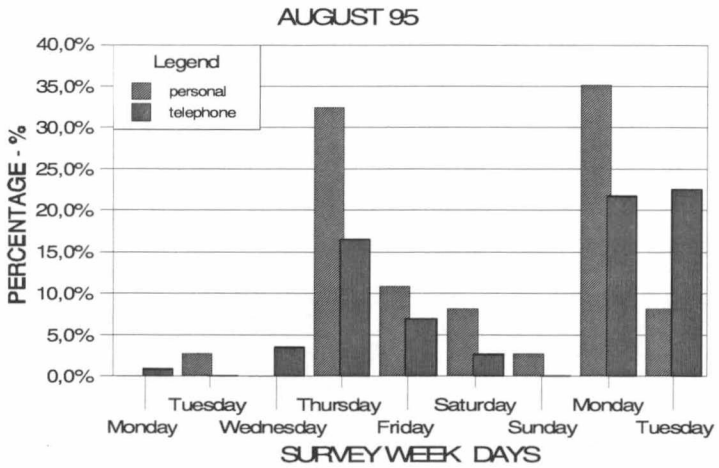
Fig. 14. Day of assigning the final code: "Temporarily absent" nonresponse.



The distribution by day, when "No one at home" codes are assigned, is given in Figure 14. Interviewers are instructed that this code should be assigned later during the week than the "Temporarily absent" households since this nonresponse component cannot be confirmed. The interviewers are to try to contact a respondent until the end of the week. Consequently, about 90 % of these codes are assigned on the last two days of the collection period – for both personal and telephone interviews.

For the "Other" nonresponse category (Figure 15), assigning codes is more spread out over the collection period. This is understandable, since this nonresponse category comprises a number of special circumstances in the household, and the code is assigned whenever they occur. The "Refusal" component is not treated here, since the corresponding data are not available yet. More work is needed to study the refusals since they are transferred to a senior interviewer for individual treatment.

Fig. 15. Day of assigning the final code: "Other" nonresponse.



As seen above, there are differences in the distribution, over the days of the week, of the assignment of final codes to nonresponding households. Also as indicated earlier, interviewers are instructed to handle different types of nonresponse differently. With these new data, it is now possible to see if the survey instructions are being followed properly. For example, interviewers are instructed that working on Sunday is up to their discretion and the data indicate that not a lot of coding is done on Sunday. The new data from the CMS will give a more complete picture of the survey and allow better monitoring of it. It is also possible to get more information about the interviewers' work and whether there are differences among Regional Offices.

Figures 16 and 17 show the frequency distribution of the duration of both types of interviews. In August 1995, the average length for a personal interview was about 17 minutes per household and the mode was 10 minutes and 30 seconds. For telephone interviews, the average length was 7 minutes and the mode was 5 minutes. This kind of information, which was difficult to obtain before CAI, can be of use: (i) to help estimate costs for LFS or supplementary survey interviews, (ii) to monitor overall costs more closely and (iii) to monitor work in the field. For example, in August 1995, some

odd cases were found. An interviewer was conducting interviews in only 17 seconds. Was it a problem with the laptop, with the CMS files or something else? Further investigation is needed.

Fig. 16. Distribution or duration of personal interviews.

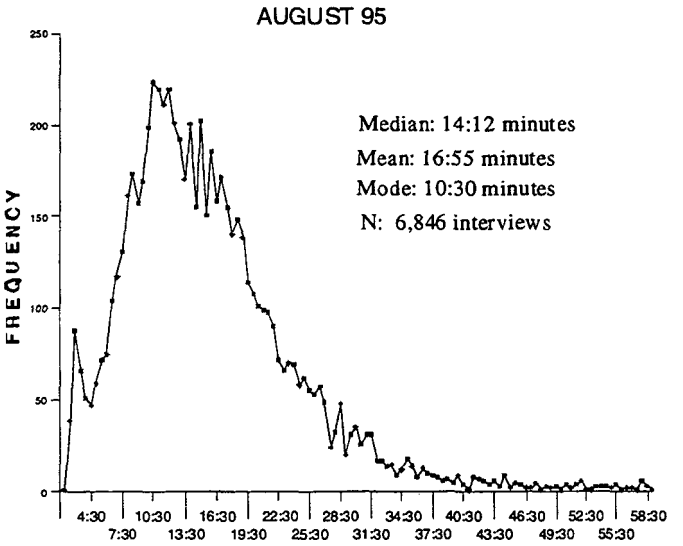
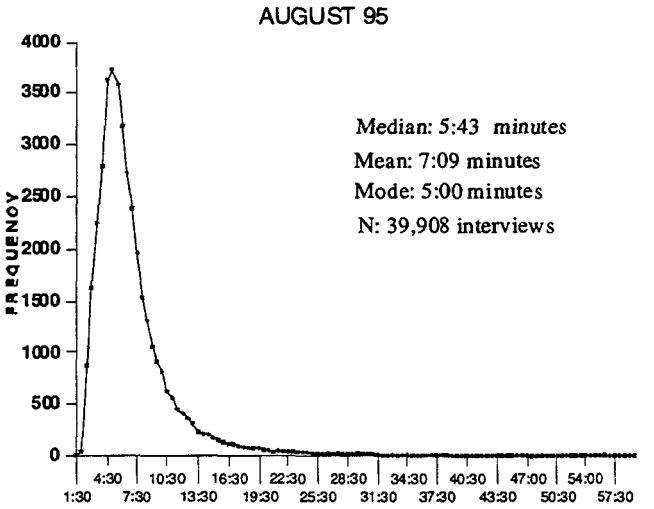


Fig. 17. Distribution of duration of telephone interviews.



5. Lessons Learned

There are five major lessons learned in the conversion of the LFS to CAI. The actual implementation of CAI took only a few months, but the effects on the quality measures of the survey lasted considerably longer. The first lesson learned was that any major change in a complicated process requires time to stabilize. For example, conversions to PAPI, as well as technical problems, which were caused by the introduction of the new technology, took about a year to disappear. Similarly, the other data quality indicators, especially the nonresponse rate, are reverting back to their pre-CAI levels only after a long adjustment period.

The second lesson is to expect the unexpected by developing contingency plans. Even though testing was done, the testing environment is never the same as the production environment, and as problems arose, people had to react quickly. For the first year, the RO field managers were simply reacting to problems as they came up. There was no time to prepare for future problems in advance. Recently, the situation has improved greatly with the stabilization of the technical problems and the PAPI conversion rate. The continuous improvement done to the CAI system is mainly responsible for this stabilization.

Thirdly, communication is essential since any relevant information has to flow smoothly between all LFS staff. Involved parties have to be informed of plans and progress, which is not an easy task in the decentralized environment of the LFS. Moreover, this information has to be communicated in a vocabulary that is easily understood.

Fourth, effective and comprehensive training of all LFS staff is crucial, especially at the interviewer level. Interviewers had to face two challenges: new technology and new working procedures. The CAI collection process requires different skills from those needed for PAPI. The interviewers, hesitant at the beginning, now generally prefer to work with CAI than with PAPI. The flexibility and the on-line editing of CAI, as well as the perceived increase in professionalism of collecting data with a computer, are features well-liked by them.

Finally, the fifth lesson is that new skills will be required to analyze the tremendous volume of new information that is becoming available from the CAI system. Just as there was a learning curve for interviewers in converting to new technology, there will also be a learning curve for survey methodologists to interpret the new data.

6. Future Work

As for the future of CAI as a data collection method, it is very promising. Based on the success of the application of CAI for the LFS, Statistics Canada adopted this mode as the standard for household surveys. There are four major areas that have been identified for development.

In addition to the information collected on labour force activities, the LFS sample is used every month to gather other important data through supplementary surveys. Consequently, a first area of development is the conversion of the LFS supplements to the CAI mode. Most of the supplementary surveys have been completely converted (or tested) to CAI. The challenge for these surveys consists of making adjustments to return to the same level of high efficiency enjoyed with PAPI.

A second major development is a new LFS questionnaire, more complex and longer, that will be implemented in 1997. This is a direct consequence of using CAI. Given the complexity of the questionnaire, it would not have been possible with PAPI. A more complex questionnaire poses new challenges when designing data processing systems (more complex edit rules, imputation methods, etc.).

A research team has been established to examine the data available from the Case Management System to develop new measures to monitor data quality, to suggest areas where improvements may be made and to find the best way to communicate the information back to the ROs. The computer programs that will be developed by this group will also be used by other household surveys since the same Case Management System is used by these surveys.

As the technology continuously improves, more efficient and effective applications are being developed and implemented. New versions of the CMS and the CAI application are expected in 1996.

References

Bench, J., Clark, C., Dufour, J. and Kaushal, R. (1994). Implementing Computer-Assisted Interviewing for the Canadian Labour Force Survey. *Proceedings of the 1994 ASA conference.*

Catlin, G. and Ingram, S. (1988). The effect of CATI on Cost and Data Quality. *Telephone Survey Methodology*, edited by R. M. Groves et al., New York: John Wiley and sons.

Dufour, J., Kaushal, R., Clark, C. and Bench, J. (1995). Converting the Labour Force Survey to Computer-Assisted Interviewing, Statistics Canada, Branch Working Paper (in progress).

Kaushal, R. and Laniel, N. (1993). Computer Assisted Interviewing: Data Quality Test. *Proceedings of the Bureau of the Census 1993, Annual Research Conference.* 513–524

Kaushal, R. and Laniel, N. (1995). Overview of Computer Assisted Interviewing: Data Quality Test. Statistics Canada, Branch Working Paper (in progress).

Simard, M. and Dufour, J. (1995). Impact of the Introduction of Computer-Assisted Interviewing as the New Labour Force Survey Data Collection Method. Internal Report, Household Survey Method Division, Statistics Canada, Ottawa.

Simard, M. and Dufour, J. (1995). The First Year of Computer-Assisted Interviewing for the Canadian Labour Force Survey. *Proceedings of the 1995 ASA conference*. To appear.

Singh, M. P., Drew, J. D., Gambino, J. G. and Mayda, F. (1990). *Methodology of the Canadian Labour Force Survey 1984-1990*. Statistics Canada Publication, Catalogue number 71–526.

NONRESPONSE: WHO RESPONDS AND WHO DOES NOT IN AN ENTERPRISE PANEL SURVEY[1]

Peter Hauptmanns
Ruhr-University
Germany

The paper describes the nonresponse analysis done in a German enterprise panel. There are three approaches for the nonresponse analysis: first, it is tried to find technical predictors that may help to identify cases that are more likely to attrition in following panel waves. Second, two surveys are presented that were done among the nonrespondents of given panel-waves, and third a comparison on key variables is done between respondents and nonrespondents of several panel waves. The result is that no systematic attrition could be identified.

Key words: Enterprise panel, prediction of nonresponse, attrition, nonresponse.

1. Introduction

Nearly every quantitative research in the social sciences has to deal with the problem of nonresponse. May it be an individual, household, organisation or enterprise survey, may it be a cross-section or a longitudinal study, the nonresponse problem hits all likewise. But there are some peculiarities in doing organisation- or enterprise surveys that lead to an increased risk of nonresponse in this kind of study.

On the other hand, we also know that nonresponse is not per-se the problem, but the question whether this nonresponse follows a random pattern or whether there is a specific nonresponse generating mechanism that may lead to a systematic bias of the survey-data. To decide this question, it is necessary to collect as much information on the nonrespondents as possible.

This paper presents what was done to collect such information in a German enterprise panel survey as an example, the so called NIFA-panel[2].

---

[1] An extended and more detailed paper (Hauptmanns 1995a) is available from the author.

[2] For more information on the NIFA-Panel see Hauptmanns 1995a.

But nevertheless it may be interesting for other research teams to see what can be done to get information on the nonrespondents and what consequences have to be drawn from this information for the validity of the survey results.

2. Gathering Information on Nonrespondents

To obtain more information about the nonrespondents and to identify a possible MD-generating mechanism in the NIFA-panel survey, three approaches were chosen:

- Some "technical" predictors for nonresponse were analysed.
- Separate surveys were carried out among the firms that did not participate in the NIFA-panel.
- Differences between respondents and nonrespondents of a specific wave are analysed by comparing the data of the two groups in the wave before (where all of them were respondents).

Technical Predictors for Nonresponse

The "technical" predictors for attrition[3] that were analysed were the response time, that means the number of days between mailing the questionnaire and receiving a response, on the one hand and the amount of item-nonresponse in a given wave on the other hand[4].

In both cases we found significant differences between the firms that do participate in the next wave and those that do not. Concerning the response time it can be shown that firms who do no longer participate in the survey in wave four needed a significantly longer time to return the questionnaire in wave three (they needed a mean of 35 working days versus 29 for the participating group). But on the other hand, more testing showed only a weak positive correlation between "response time" as independent and "drop-out in the next wave" (eta = .28).

Very similar results were obtained by comparing the participating firms with the drop-outs concerning the item nonresponse in the wave before the actual one. The basic idea behind this is, that if firms make the experience that filling out the questionnaire is either hard and costly or that it is not

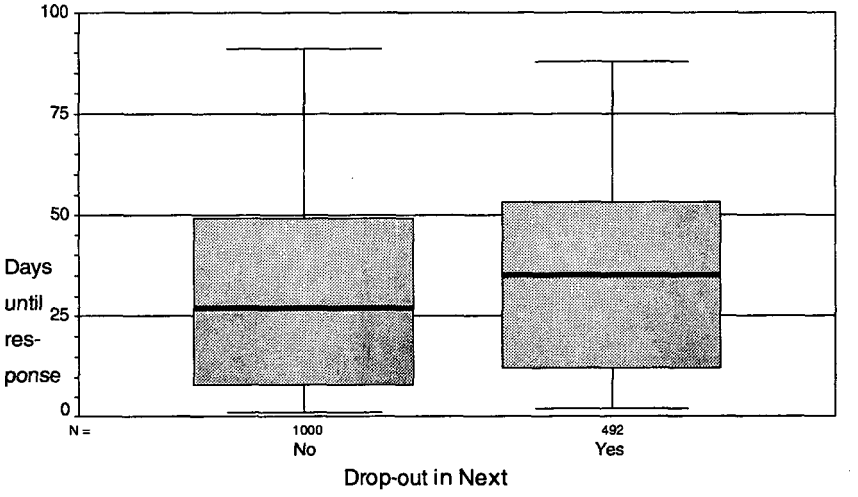
---

[3] Attrition is meant in this paper as "refusal to participate in the survey after doing so for at least one panel wave before". Real mortality, which means the firms do no longer exist in the population, is excluded from this analysis.

[4] The tests documented in this paper show the results for the analysis of third vs. Fourth wave, as in the first and second wave the returning questionnaires were given a number according to the order of receivement, but that is no valid indicator for the response time. When we changed the institute that carries out the survey after the second wave, the date itself was saved.

adequate for them, this might lead to more item-nonresponse. Keeping this experience in mind the probability for a refusal to participate again should decrease[5]. Our analysis indeed showed some proof of this hypothethis: among the group of firms which did not participate in wave four, the rate of item-nonresponse was significantly higher than among the firms which did stay in the panel (a mean of 22.5 missing items in the one group versus 18.9 missing items in the other). Again we find an eta value of .29 between these two variables.

Fig. 1. Response time and attrition.



Although we do not want to construct any causal relations between response time or item- nonresponse and panel attrition, it might be a good idea to invest some more resources in the following waves in the firms that showed one or both of these attributes (long response-time and high item-nonresponse rate).

We also tested the idea that there might be a correlation between response time and item-nonresponse (and thus maybe an important interaction effect for the analysis of unit attrition), but this proved to be wrong. A logistic model with "attrition" as dependent and item-nonresponse, response time and the interaction term as covariates showed no significant effect of the interaction term[6]. The non-existence of a correlation between these two effects can easily be derived from the following plot[7]:

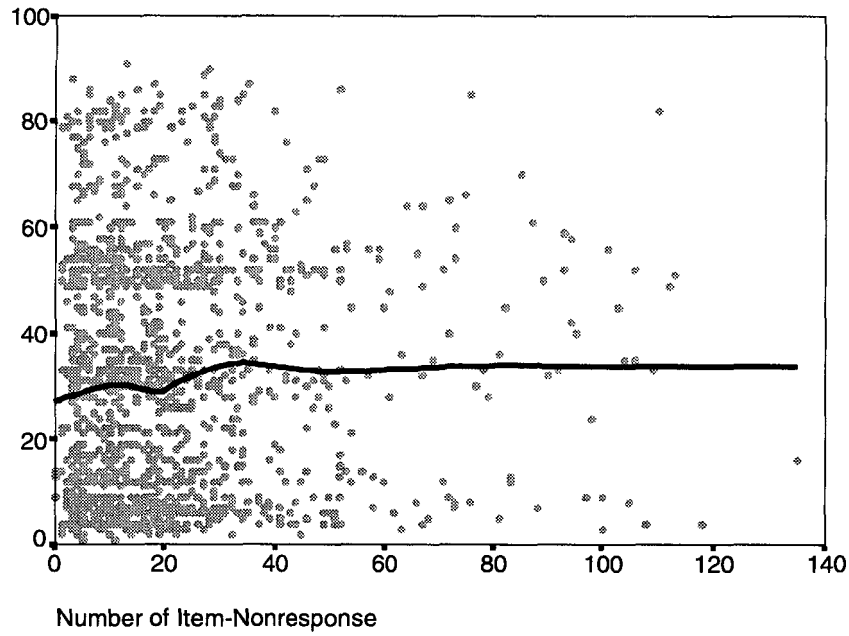
---

[5] This analysis was done after the Helsinki-Workshop. We did not intend to test this. Credits and thanks for this idea go to Mick Couper (JPSM) and other participants of the Workshop.

[6] Both the main effects had a significant B-value and were entered in the regression; nevertheless the model did not fit very well, so we assume that attrition is not predictable by using those technical predictors alone).

[7] The fit-line is a lowess smoother.

Fig. 2. Scatterplot response-time vs. amount of item-nonresponse.



In another test we analysed the correlation between response time and some of the main theoretical variables of the survey. A test of the willingness to participate in the survey showed that there is indeed a positive correlation between the speed of response and the technical equipment implemented in the firm (eta = .36); that means firms which did not reply before the second reminder usually have less CIM-technology implemented than early respondents – which simply means they do not look at the survey as adequate or salient for them. A t-test of early versus late respondents[8] showed difference in the means of the variable "number of CIM technologies implemented". This leads to the conclusion that nonrespondents will have less or even no CIM-technology implemented, or, in reverse, firms that did participate in the survey will have more CIM-technology implemented than nonrespondents. This hypothesis was confirmed by the nonresponse-study after the first wave in 1991 – but no longer by the nonresponse study of 1994. The next chapter will show the results of these studies in more detail.

[8] This analysis was carried out after the first wave. As mentioned above, in the first two waves the exact date of the receivement of the returning questionnaire was not saved. We used the order of incoming instead, and to minimize the risk of misinterpretation, we dichotomized this variable to early and late respondents, where early respondents means the questionnaire was received among the first 75 % that answered, and late respondent means it was received among the last 25 %.

The Nonresponse Studies

The 1991 survey After the first panel wave in 1991 a separate nonresponse study was carried out to obtain deeper knowledge about the structure of the nonrespondents. It was our plan to find out whether any particular variable or set of variables increases respectively decreases the probability of participation in the NIFA-survey.

To obtain this information a random sample of some 5 % (= 135 firms) was drawn out of the nonrespondents. These firms received a postcard with four questions: whether the firm agrees in the affiliation to the mechanical engineering industry, whether the firm actually produces machines or machine tools at the respective address, whether the firm has any CIM-technologies implemented and why the firm did not participate in the survey. Moreover we had the information about firm size and industry affiliation from the FES register.

We were aware that these four questions alone would not be sufficient to obtain a broader knowledge on the nonrespondents. On the other hand it is very difficult to convince a group of nonrespondents to a survey to participate in a follow-up study. So our thought was that it might be more valuable to get only few information, but from as much as possible participants than to get more information from less firms (this was the – unintended – result of the 1994 nonresponse study).

Several reminders were given by phone, so that at last 127 out of the 135 firms participated in this nonresponse study (two more letters were returned with the note "address unknown"). Firm size and industry affiliation were added to the data from the FES register.

The question why the firms did not participate in the NIFA-survey caused no surprise: 63 % gave a lack of time resp. the length of the questionnaire as reason. 26 % said they would refuse to answer questionnaires in general, 12 % saw themselves not as a mechanical engineering enterprise, and some 4 % gave other reasons (multiple responses were allowed).

The distribution of firm size showed – as expected – an over-representation of small firms. 48 % of the participants in this sample had 20–50 employees, versus 43 % in the population. As small firms were underrepresented in the main survey (compared to the FES register data as population) this had to occur if the sampling was correct.

The analysis of the use of CIM-technologies confirmed the results mentioned above: whereas in the main survey only 11 % of the firms said not to use any CIM-technology 26 % in the nonrespondents study gave this answer ($t = 2.98$; $df = 109.8$; $p = .002$). An ANOVA-model made clear that not the firm size but the use of CIM-technology was the decisive factor for the willingness to participate in the NIFA-survey. Though respondents and nonrespondents showed significant differences on both variables, in the

multivariate analysis the technology variable alone had a significant effect on the willingness to respond ($F^{1}_{2649} = 11.809$, $p = .002$). The factor "firm size" (categorised) on the other hand showed no significant effect ($F^{5}_{2649} = .739$, $p = .594$).

The 1994 survey After the fourth wave another non-response study was carried out. This time a 5 %-sample out of the group of firms that had never taken part in one of the four panel waves was drawn. These 135 firms (100 in west and 35 in east-Germany) received a questionnaire containing 9 questions. These questions covered the following key-variables of the NIFA-survey. Again the variables firm size and industry affiliation could be added to these variables[9] .

Based on the experience of the first, small nonresponse survey we had the assumption that our population might be split into two segments:

- one that consists of modern firms that used CIM-technology to a certain extent, that faced a severe economic competition and is thus confronted with high demands for an organisational and production flexibility and that judges the future economic developments as not so good;
- and one that consists mainly of very small firms, producing for market niches without great competition, without a hard pressure for flexible technology or organisation from outside, e.g. by customers or competitors, and that judge the future economic development of the firm more stable and satisfying.

These two groups would be differentiated according to the widely accepted theoretical discussion in industrial sociology or organisational theory. Our assumptions was, that, if this differentiation of the population was true, we should find more firms of the first group among the respondents and more of the second group among the nonrespondents. That hypothesis was set up with the assumption in mind that firms, who produce in stable environmental settings without a need for organisational and technological changes would see no advantage from resp. no use in participating in a survey that covers mainly these items of change.

Thus we thought we should find some significant differences between the two groups on the measured key variables. Surprisingly, this assumption was wrong. We did not find any differences between the respondents and the nonrespondents group on any of the measured variables.

To test our hypothesis we first looked at the firm size and the use of CIM-technology, the two variables that proved different between the two groups in the 1991 survey. Then we compared the means of all variables available in both groups. None of the differences gave any positive proof for

---

[9] At least for the west-German firms. These informations are not yet part of the FES-register for east-German firms.

our assumption of a split, homogeneously segmented population. Even multivariate analysis (Logit and CART-analysis) with "response" as dependent and all of the above as independent variables did not produce any model with a satisfying fit.

Thus we have to reject this hypothesis (what we do happily, of course) and accept that the nonresponse is not clustered in a specific segment of the population, instead. Concerning the title question, who responds and who does not in an enterprise panel survey, we have no definite answer – there is not a specific group that responds and another that does not. Nonresponse seems to occur completely at random in this survey. On the other hand, we do not believe in people flipping coins to decide whether to participate in a survey or not. But what we at least know is that the MD-generating mechanism is not correlated to any of the variables we are especially interested in.

This result differs from the one described above for the 1991 nonresponse survey. This may be due to several reasons: on the one hand, it may be that the information gathered in 1991 was not adequate and sufficient to come to a satisfying and valid conclusion. On the other hand – and we take that as more likely – there may have been a shift in the population: in the years of economic crisis a lot of firms ceased to exist, larger firms were broken up into several small units, the idea of "lean production" became prominent. It is a reasonable assumption that most of all small firms that are not equipped with modern technology and therefore had no potential to meet the demands for a less costly and still more productive manufacturing process organisation were the first to close. This means that the more dramatic changes in the population, the higher attrition, might have taken place in the group of nonrespondents of the first wave. Thus the results given above have two faces – the happy face, because we were assured that our sample is a fairly good representation of the present population, and the sad face, because the survey did not cover the most dynamic and dramatic changes that might have taken place in the population.

Comparison within / between Panel Waves

In another step of analysis we tested whether we would find any differences between firms that stop participating in the panel after two or three waves versus such that continue to participate over all waves.

We made a first test in 1991 by comparing the firms that participated in the first waves main survey with the ones that only took part in the basic survey (cf. Flimm and Saurwein 1992). These two groups differ neither on firm size nor on industry affiliation (hardly any other interesting variables were measured in the basic survey of wave one). A significant effect of the "technology"-variables is not detectable either, neither for the question

whether they use CIM-technology at all nor for the number of CIM-components implemented.

A comparable analysis was carried out for the group of firms that did participate in the first two waves of the NIFA-panel but not in the third. Therefore two groups were constructed in the data set for the first two waves: "respondents 93" and "nonrespondents 93". With several different statistical tests we tried to identify significant differences between these two groups for a number of variables. But none was found – None of the analysed variables showed a significantly difference in mean/median or distribution between respondents and nonrespondents. So we may assume that panel attrition is not caused by an explicit MD-generating mechanism but occurs more or less at random.

To confirm this we tried to identify variables which may have a significant impact on the drop-out for one wave. We first applied a classification tree with the software CART (cf. Breiman et al. 1984), an ideal tool for data exploration. The model CART generated was then tested in a logit analysis several times with modifications. Even with this procedure we could not identify a model with an adequate goodness of fit for "panel participation" as dependent variable. The data set analysed did not include any variable that had a significant effect on the willingness to participate in the panel continuously resp. to drop out of the panel. Thus, we can be rather sure that attrition in the NIFA-panel is based on random effects only. This result could also be confirmed with a micro-simulation approach for drop-outs between the first and second wave (Niggemann 1993).

We did not yet repeat this analysis for the fourth wave in 1994, as we do not have any reason for the assumption that these results are no longer valid in that wave. We will, however, keep on checking this with next waves data.

3. Conclusions

This paper was started with the question "Who responds and who does not in a firm panel survey". The first and most important conclusion from our analysis is, that this question cannot be answered in its original sense. We saw that respondents and nonrespondents to this survey do not form two separate and internal homogenous groups that can be grouped among several key variables. A – positive – result from the analysis is that both groups do not differ on these variables significantly and therefore we can assume that we have to deal with a random process that generates the nonresponse.

This does, on the other hand, not mean that this nonresponse is completely ignorable and we would not have to care about it any more. As

said above, the non existing differences in 1994 – after finding differences between the two groups in 1991 – may be the result of a dynamically changing population. The main focus and intention of a panel survey is to identify such developments. Thus we have to learn that we must have a keen eye on the nonresponse-group in the future to be sure that important developments in the population will be covered by the panel survey.

Every survey has to deal with a more or less high rate of nonresponse, that is undoubted. Nevertheless, gathering information about the nonrespondents, doing nonresponse surveys, identifying nonresponse mechanism and so on are only the second best choice. The primary target has to be to avoid nonresponse with all possible means. That becomes especially relevant for panel studies. Not only may a high nonresponse rate lead to misrepresentation and skewed distributions in the time series, but panel attrition can even lead to a situation where complex multivariate analysis (which are mostly the most important mean of testing hypothesis, finding and analysing dynamical developments and thus producing new knowledge about the research topic) become impossible because of too few cases left in the panel. Therefore the importance of "panel care" or maintenance cannot be overestimated.

References

Esser, H. (1986). Können Befragte lügen? Zum Konzept des "wahren Wertes" im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. 38, 314–336.

Hauptmanns, P. (1995). Methodological Experiences and Problems of a Firm Panel Survey, 267–279 in: European Commission (ed.): *Techniques and Uses of Enterprise Panels*. Office for Official Publications of the European Commission, Luxembourg.

Hauptmanns, P. (1995a). Nonresponse in an enterprise panel survey. *Bochum: SFB 187 Working Paper.*

Hox, J. J. and De Leeuw, E. D. (1994). A comparison of nonresponse in mail, telephone and face-to-face surveys. *Quality & Quantity*, 28: 329–344.

Kruskal, W. and Mosteller, F. (1979). Representative Sampling I: Non-Scientific Literature. *International Statistical Review*, 47: 13–24.

Lepkowsky, J. M. (1989). Treatment of Wave Nonresponse in Panel Surveys, 348–374 in Kasprzyk, Daniel, Duncan, Greg, Kalton, Graham and Singh, M. P. (eds.). *Panel Surveys*. Wiley & Sons. New York.

Niggemann, H. (1993). Die Anpassung von Restriktionen an Mikrodaten mit dem statistischen Mikrosimulationsmodell MICSIM: Ein Konzept zur Analyse von Ausfallprozessen im NIFA-Panel. *SFB 187 Arbeitspapier, Bochum.*

Schnell, R. (1993). Die Homogenität sozialer Kategorien als Voraussetzung für "Repräsentativität" und Gewichtungsverfahren. *Zeitschrift für Soziologie*, 22, 16–32.

NONRESPONSE AND ILL-BEING IN THE FINNISH SURVEY ON LIVING CONDITIONS

Markku Heiskanen and Seppo Laaksonen
Statistical Methods and Research Branch
Statistics Finland

Nonresponse was substantially higher in the 1994 Living Conditions Survey (SLC) than in the previous two surveys. The most important aspects of this are: (i) The SLC was combined to another survey, the Income Distribution Survey (IDS), (ii) field work was not as effectively controlled in 1994 as earlier, (iii) the economic recession of the 1990s was still continuing in 1994, and this may have influenced the survey climate. The impact of these factors on the response rate cannot be fully justified, but we have analysed nonresponse in different population groups, especially in those suffering from ill-being (poverty). The data set of the SLC straightforwardly offers some register information for this purpose. From the point of view of the nonresponse analysis, we had an additional opportunity to use specific auxiliary information about people living in a "poverty trap". Results show that experienced unemployment in the family does not essentially increase nonresponse, but if the household of a respondent has received living allowance, nonresponse among men was considerably higher. Men seem to be more sensitive to nonresponse in "poverty trap" groups than women, particularly if they live alone. The nonresponse analysis is made both descriptively and using logistic regression with a number of explanatory variables. The estimated response probabilities are also utilised in the methods of adjustment for nonresponse, the effects of these are compared with other estimates.

Key words: Logistic regression, poverty, reweighting.

1. Introduction

The Finnish survey on living conditions (SLC) conducted by Statistics Finland has been carried out three times, in 1978, 1986 and 1994. The questionnaire of the first survey differs essentially from that of the two latter ones. In addition some new areas were added in 1994 compared with the 1986 survey. The new questions of the 1994 survey concern, for example, subjective feelings and values.

The main topics of the 1994 survey were:

- household/family structure,
- income and wealth,
- living conditions in childhood,
- residence and housing conditions,
- social activity and leisure activities,
- contacts and relations to relatives and friends,
- education and training,
- work and working relations,
- health,
- feelings of well-being and of oneself,
- security and safety.

The aim of the SLC is to produce a cross-sectional description of well- and ill-being of the Finnish citizens so that the various aspects of living conditions would be included in the same survey. The SLC is based on the Nordic welfare study tradition (welfare as resources) and also on the OECD Social Indicator Programme. Many questions are comparable with Nordic living conditions surveys, and comparative reports of the data have also been published (e.g. Vogel 1991, a more detailed study of health inequalities in Finland, Norway and Sweden, see Lahelma et al. 1993).

The SLC data have been used for many purposes in Finland. One is the descriptive statistics of the life of citizens compiled mostly by Statistics Finland. The data have also been used widely in academic studies. Dissertations on poverty and deprivation in a welfare state (Heikkilä 1990), on working at home (Salmi 1991) and those on working environment are examples of different areas of social study. It should be noted that many of these studies were focused on marginal groups, and these results can be very sensitive to bias due to nonresponse or measurement errors, in particular. Correspondingly, all attempts to improve the quality of the data are highly motivated.

The SLC has been greatly appreciated at Statistics Finland because of its many social and scientific connections. One crucial requirement for the 1994 survey was to maintain the high quality of the data and the comparability of many question sets with the previous survey and with the Nordic results (Ahola et al. 1995). Because of the nature of the SCL as a means of locating and describing social problems at different social strata, especially the representativeness of the population groups suffering from ill-being was considered very important. It is known that nonresponse often tends to increase among groups suffering from ill-being, thus the high response rate and the ignorability of the response mechanism (see Little and Rubin 1987) have been the requirements of the fieldwork.

However, it is to be expected that these targets will not be ideally obtained in the SLC or any other large-scale surveys. This paper aims at analysing the response rates of the 1994 SLC, and at comparing these figures with the previous surveys. Both descriptive and modelling approach are used to analyse the characteristics of response vs. nonresponse. Finally, attempts are made to adjust for the bias due to nonresponse by reweighting and to compare these estimates with those based on less effective adjustment methods.

2. Descriptive Analysis of Respondents

There are some differences between the 1986 and the 1994 surveys. First, the latter was conducted as a computer-assisted personal interview. This feature has no direct connection to nonresponse although the transition to the CAPI method was feared to cause practical problems among interviewers.

The second difference is more severe as regards nonresponse: the SLC was connected to the Income Distribution Survey (IDS) to make possible to finance a sufficiently large sample size. The response burden of the target person/household (the IDS uses a household as the survey unit, the SLC a specific member of the household, called a target person, respectively) increased – they were informed of both surveys in advance. The interviewing averaged 80 minutes, of which 55 minutes were used for the SLC.

Also the sampling method varied. A systematic random sample of Finns aged 15 or over was drawn from the Central Population Register for the 1986 SLC. The 1994 SLC consists of two panels of the IDS, the former being drawn as a multistage sample from the State Taxation Register and interviewed for the first time in 1993, and the latter from the Central Population Register in order that certain quotas of different social strata would be attained.

As the nonresponse of the 1993 panel was included as nonresponse of the SLC – no attempts were made to contact the 1993 nonrespondents in 1994, and because of the larger response burden and the different target populations, nonresponse was expected to rise from that of the 1986 survey. There were also signs of deterioration of the survey climate due to the severe economic recession.

The sample of the SLC thus consisted of two panels of the IDS totalling 12,093 persons, and 11,843 persons without overcoverage. The number of accepted interviews in the data set amounted to 8,650.

The nonresponse rate in the 1994 SLC increased to 27 per cent, whereas it was only 13 per cent in the 1986 survey. Although some of the nonresponse cases might be classified as "technical", e.g. part of the non-

response cases of the 1993 panel might be shifted to overcoverage cases in the second wave of the panel, it is a fact that nonresponse was a worse problem in the 1994 SLC. Although a slight increase in overcoverage and noncontacting occurred, the main reason for the increased nonresponse are refusals.

Table 1. Response rates in Finnish Living Condition Surveys.

Year	Response rate, (%), Total	Response rate, (%), Men	Refusal, (%), Total	Sample size, (n), Total	Overcoverage (%)
1978	84.9	84.9	6.2	2,971	1.4
1986	86.9	96.7	9.3	13,876	1.6
1994	73.0	72.1	21.9	11,843	2.1

The response rate of the "pure" IDS is considerably higher than that of the 1994 SLC. For example, for 1991 the rate is 83.4 per cent, for 1992, 84.3 per cent and for 1993, 84.9 per cent, whereas a mere 76.5 per cent for the first wave of the panel in 1994. Thus, it seems that combining of two surveys has given rise to increased nonresponse. There are similar findings from the 1980s. The increase in nonresponse is primarily caused by refusals: the refusal rate is on average about 12 per cent during the period 1990–1993, whereas in 1994 it is 19 per cent.

Somewhat contradictory to the response burden argument is that even in the pilot study of the autumn 1993, consisting of 1,000 persons, the nonresponse rate was fairly high, at 24 per cent. The IDS was not included in this survey. The average refusal rate of the pilot survey rose to over 17 per cent.

The survey designs of the SLC and the IDS are not identical. The combined survey was linked together with three CAPI modules or questionnaires. The interviewer used a CAPI-menu to select the modules. The interviewing order was following: firstly, the household background module, secondly, the SLC and finally the IDS. To reduce nonresponse the interviewers were given the possibility to make the IDS interview by telephone if the respondent was not willing to participate in a face to face survey. They were also allowed in some cases to substitute the proper target person by another member of the household, which was not possible for the SLC. Such cases in the IDS are mainly classified as refusals in the SLC.

There are some external factors contributing to the final response rate. During the 1986 SLC the high response rate was one of the main goals to be set to the field work. Regional interviewer groups were built up to take care of nonresponse, among other things. In the 1994 survey the emphasis in the field organisation was placed on taking care of the functioning of the new computer-assisted working environment of the interviewers.

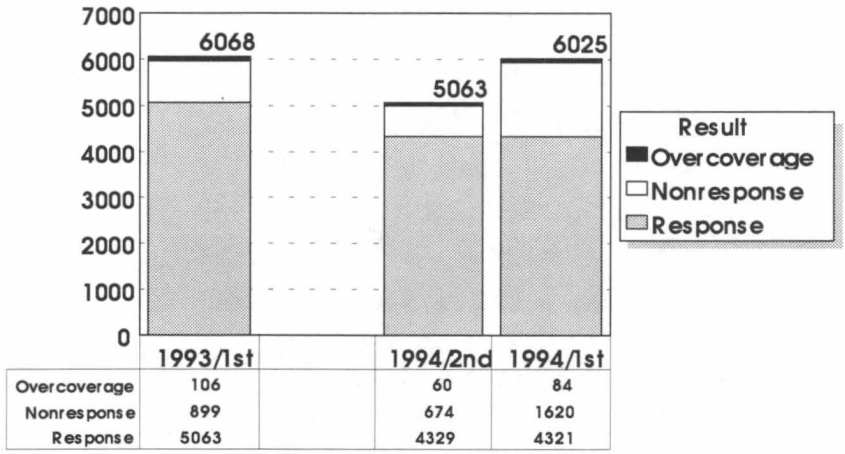
Nonresponse may have increased to some extent due to a subsample of 1,001 persons from the Greater Helsinki region. In this region nonresponse

has traditionally been higher than in other parts of the country. The absence of 15–16 year-old target persons in the 1993 panel also increased the unit nonresponse somewhat. In addition, the economic recession and high unemployment impaired the 'survey climate.' Too much emphasis should not be put to this argument though, because in the Finnish Health Security Survey in 1995 the response rate was nearly 90 per cent. The duration of the field work was also shorter in 1994 compared with 1986. Thus it is possible that the number of small details in the survey design as a whole have given rise to a somewhat alarming nonresponse.

Figure 1 illustrates the increase in nonresponse in the first wave of the 1994 panel. It is twice as high as for the first wave of 1993. The nonresponse rate of the first 1993 wave and the second 1994 wave together is of the same size than that of the first 1994 wave.

The response rate in 1994 is, however, quite even or at least traditional in different population groups (gender and age group – Figure 2). Nonresponse seems to be "normal" (although at a higher level) also by marital status, incomes and education: people who are engaged in better positions have been more willing to co-operate.

Fig. 1. Response and nonresponse in different waves in 1994 (persons).



	1993/1st	1994/2nd	1994/1st
Overcoverage	106	60	84
Nonresponse	899	674	1620
Response	5063	4329	4321

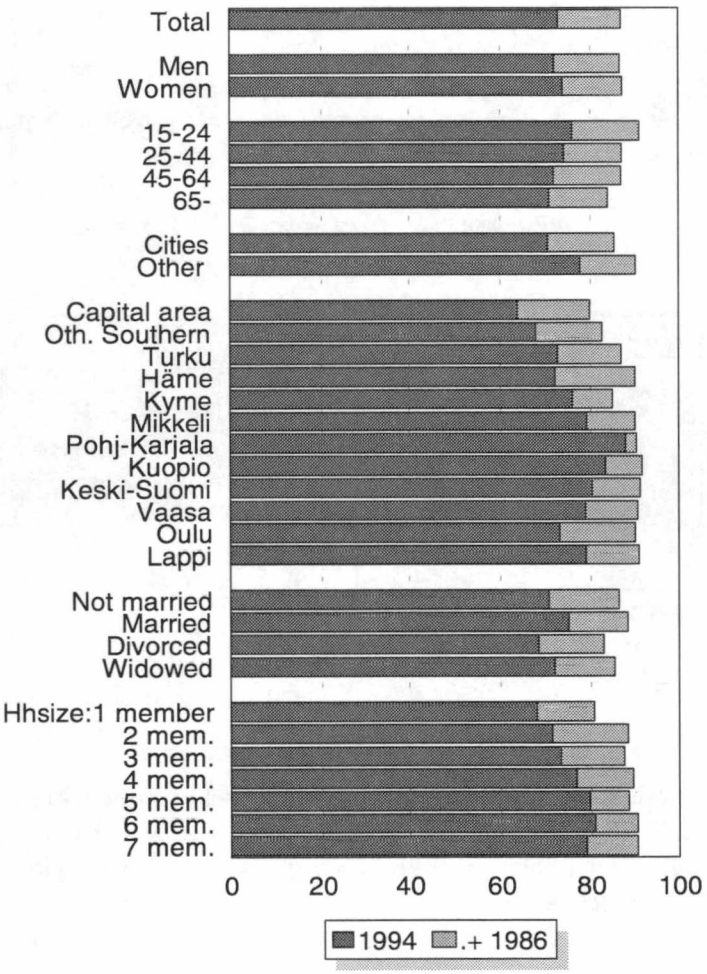
There are, however, differences in regional response rate: in the eastern part of Finland the decline in response is smaller than in the rest of Finland; it is as low as 2.4 percentage points in one province. Irrespective of the regional differences, we are inclined to think that due to the different sorts of reasons for nonresponse, the nonrespondents are fairly evenly distributed in different population groups.

Because the SLC is nevertheless a social survey, which tries to expose social problems and find out population groups that are doing non-well or

ill, we tried to find information on people suffering from economic problems. In Finland, we have many administrative data sets, which Statistics Finland is allowed to use for statistical purposes.

Information called "recession trap variables" obtained from the register sources of the National Research and Development Centre for Welfare and Health (STAKES) was connected to the SLC data set. The data between the registers and survey data sets may be linked via the personal identification number given to all Finns. The register contains information on people whose family has at least some time during 1992 had economic problems. It is supposed that this "registered ill-being" has effects on the life of these persons also two years later, in 1994.

Fig. 2. Response rate in the 1986 and 1994 SLC.



The six variables of the STAKES register were connected to the SLC data set. These are unemployment (of the target person or his/her spouse), low

income (the income of the family lower than half of the median income of the households), over-indebted (a household must have turned to social bid for loans), living allowance (living support is paid to the family because of insufficient incomes), rescheduling of private persons debts (the authorities as an intermediary) and person-based bankruptcy.

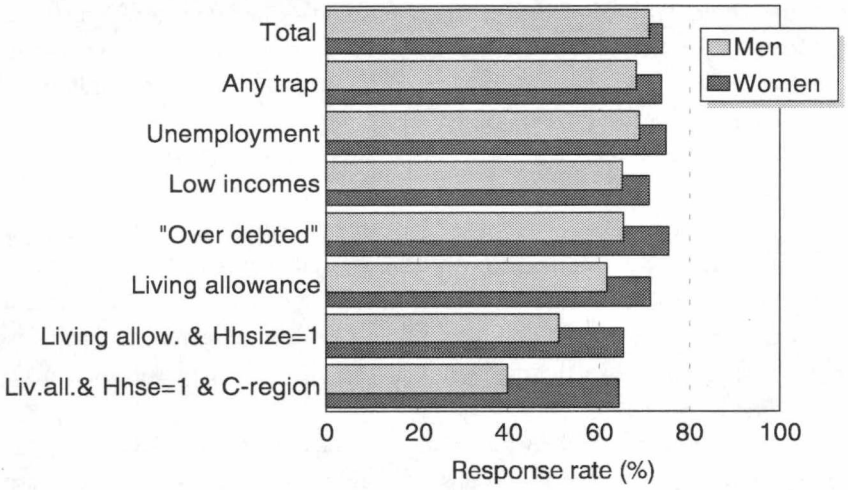
Thirty-one per cent of the sample families fell into the group "at least one trap" during 1992. Twenty-three per cent suffered from unemployment. Living allowance was received by 9 per cent, 8 per cent of the sample earned less than half of the median income. These variables might be classified as conventional poverty measures. Rescheduling of debts (0.4 %) and failure (or bankruptcy) (0.1 %) is less common and is directed to different population groups than the conventional measures, so they were excluded from further analysis.

Figure 3 shows that for women, if the family belonged to the recession trap group (or any of its subgroups), the response rate does not decrease considerably. In fact, in groups of unemployed and over-indebted women the response rate is even slightly higher than that at the total level.

The situation of men is different: their response rate decreases when they move to groups having more severe economic problems. Unemployment is so common in the 90s (and regarded as socially tolerated) in different population groups that it does not drop the response rate heavily, although a small decline occurs. But receiving living allowance makes the response rate drop by nearly 10 percentage points.

We might ask what happens if the target person is a man, receiving living allowance but not family support? In that group the response rate is still lower, 51 per cent (number of observations, n = 168). The response rate for male respondents getting social support and living alone in the Helsinki region (note: Helsinki is the only larger urban region in Finland) is still lower, 40 percentage points (n = 45). It is possible that the response rate might be even lower than that among the ones still worse off, e.g. suffering from mental problems or alcohol addiction. People having no residence or living in institutions were not included in the sampling frame but they might be classified into one of the worst groups and these persons account for 1.5 per cent of the inhabitants.

Fig. 3. Response rate of different "poverty traps" by gender.



Reason for the increase in nonresponse was noncontacting. Refusals did not increase as poverty increased.

3. Logistic Models for Nonresponse

The above descriptive analysis gives useful results, but the following modelling approach is more compact and multi-dimensional. The models may also be used in further adjustments for nonresponse as we show in Section 4. Since three types of data sets are available, the three various models with similar principles were constructed:

- Model A, for the first wave of the panel from 1993 to 1994,
- Model B for the second wave of the panel from 1993 to 1994, excluding nonresponse of the first wave,
- Model C for the cross-sectional survey of 1994.

Model B is dependent on Model A, but Model C is independent of both of them. We have no clear expectations on the dependency of Models A and B. Nonresponse of Model A is due only to the Income Distribution Survey (IDS), but that of Model B both to the IDS and the SLC. The nonresponse rates of Model B are lower than those of Model A due to exclusion. Correspondingly, there should be some differences in nonresponse behaviour both due to different surveys and different waves.

The explanatory variables are first classified into three groups:

(i) traditional nonresponse background variables obtained from the Central Population Register. These variables are linked together with the sample. The reliability of the background variables is high (e. g. in 94 per

cent of the cases the marital status in interviewing corresponded to the register information)

(ii) ordinary register variables used by Statistics Finland, which are added to the data set in many studies: e.g. education and incomes.

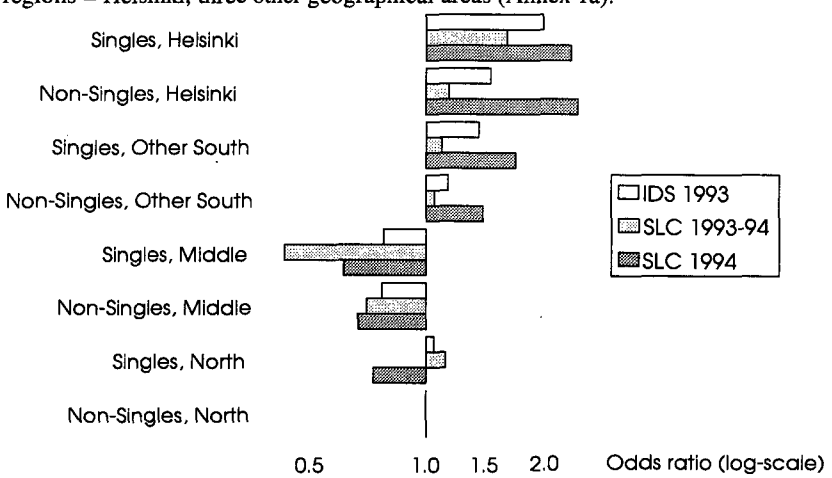
(iii) additional variables, which here are called the recession trap variables.

A number of alternative models were tested in order to find the best explanatory variables and their classifications. However, we wanted to use the same explanatory variables in all the three models in order to better compare the estimates. The list of variables is presented in Annex 1a.

The parameter estimates of all the three logistic models are presented in Annex 1b. In addition, we consider some special features of each model using estimated odds ratios in Figures 4a, 4b, and 4c.

Fig. 4a. Estimated odds ratios for nonresponse by region and household size.

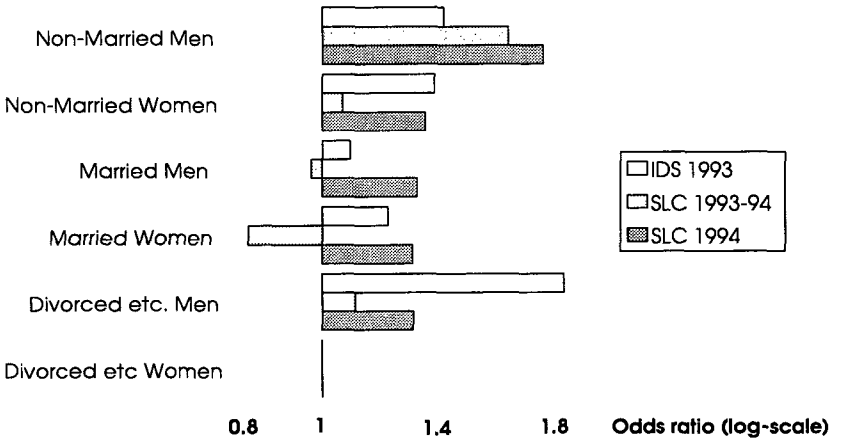
Notation: single = one-person household, non-single = other households; regions = Helsinki, three other geographical areas (Annex 1a).



IDS = Income Distribution Survey; SLC = Living Conditions Survey

Figure 4a shows that the sampled people of the Middle Finland have responded better compared with the Helsinki area, in particular. This regional feature is similar to other Finnish surveys. On the other hand, singles (one person households) themselves are not as problematic as we expected. As regards this point, it should be noted that the logistic model also consists of the specific variable 'Marital Status' which is very important as Figure 4b implies.

Fig. 4b. Estimated odds ratios for nonresponse by marital status and gender.



IDS = Income Distribution Survey; SLC = Living Conditions Survey

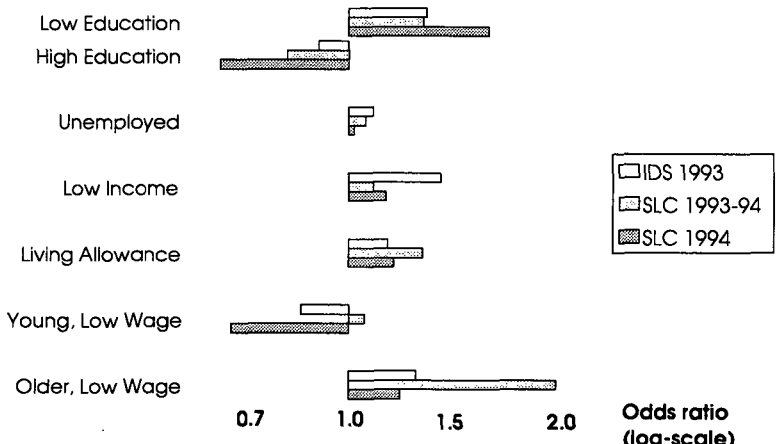
Figure 4b proves that nonresponse for both classes of men without spouses, (that is, non-married or never-married, and divorced and widowed men) is higher than for similar women. This is not surprising, but instead the fact that the lowest average nonresponse rates are estimated for divorced and widowed women. In addition, the nonresponse rate is far away from that of the corresponding men. Further, it is interesting that the lowest estimated nonresponse rates and odds ratios are obtained for married women and men of the second wave of the panel from 1993 to 1994, thus concerning only the respondents of the first wave of the IDS. The additional burden due to the second wave of SLC thus was not so hard for married people.

Figure 4c gives estimated odds ratios for variables of the second and third group mentioned above. They show that the effects of the level of education or wages on responding are more significant than those of our recession trap variables. The difference between low education (no occupational qualification) and those who have attained at least the lowest academic degree has increased in the 1994 SLC compared to the other waves.

Younger people with low wages (aged under 29 & income less than lowest quartile of taxable income) were very willing to participate in the 1994 SLC. The age classification covers the majority of at least the youngest population groups of the sample and is not strongly determined by incomes. E.g. at the age bracket 15–24 years the response rate is 3.1 percentage points higher than on average. People older than 28 years and having low incomes may, however, be classified as a true poverty group.

The estimates of the three different data sets are in line with each other in most cases. This means that they do not differ dramatically from each other and that the effects of the 1993–94 SLC are between the estimates of the other data sets.

Fig. 4c. Estimated odds ratios for nonresponse by some social status variables. The reference group with odds ratio equal to 1 is a missing class, e.g. 'employed' for 'unemployed'.



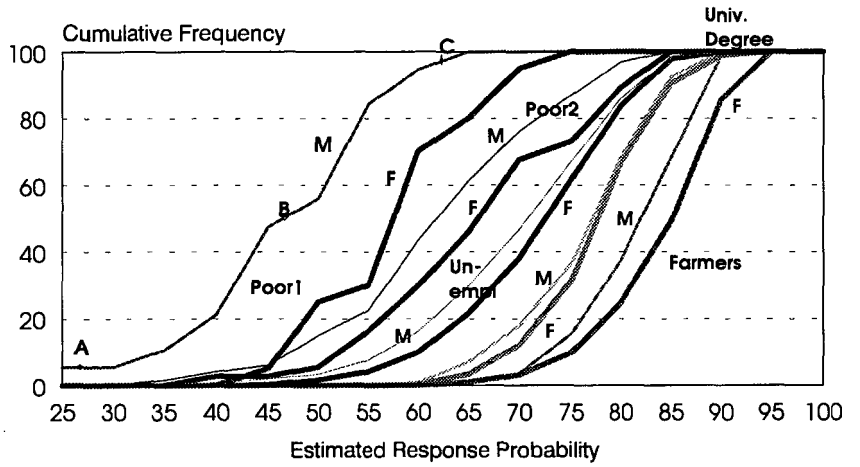
IDS = Income Distribution Survey; SLC = Living Conditions Survey

Logistic regression can be used in many ways. In Section 4 we consider its utilisation in adjustments. The estimated response probabilities can be used to illustrate differences in responding of various sub-groups of a sample by computing these predicted probabilities for each respondent group. As a result, we have theoretical probabilities via which we can construct distributions, for example. We here induced the cumulative frequencies for some interesting groups. In Figure 5 certain population groups have been chosen.

The group Poor1 (single poor people in Helsinki) is not large, but its estimated response probabilities are the lowest in our data set. On the contrary, farmers have traditionally been very willing to respond in surveys conducted in Finland.

We can observe differences in responding between men and women. The gender difference is broadest for poor people of type 1, fairly broad even for farmers, whereas for people with university degree it is small. This finding gives evidence to our hypothesis on the descriptive analysis that was presented in Section 2.

Fig. 5. Cumulative frequencies of estimated response probabilities for some population groups by gender.



Notation: Poor1 = received living allowance, single, Helsinki;
Poor2 = received living allowance, single, non-Helsinki;
M = Male (thin curve), F = Female (thick curve)
Some specific points:
A = poor1 & 52 years & 1st decile wage & low income & low educ. & manual worker & non-married
B = poor1 & 51 years & 3rd decile wage & unemployed & low educ. & non-man. worker & divorced
C = poor1 & 24 years & 6th decile wage & unemployed & medium educ. & non-man.worker & non-married

The distributional characteristics in Figure 5 can also be interpreted. The specific three points A, B and C of the first curve illustrate the multi-dimensional characteristics of these individual respondents (note: not all the variables are included). More generally, we can see, for example, that a certain kind of minimum response rate for the male poor people of type 1 is about 25 per cent, whereas the maximum level is close on 60 per cent. We also see that the worst group of farmers responds in 65 per cent of the cases, whereas their best group replies in 95 per cent of the cases (women). It is interesting that the distribution is fairly wide for poor people of type 2, but not particularly narrow as regards unemployed people.

4. Effects on Estimates with and without Adjustments

Nonresponse should be dealt with throughout all the phases of a survey and moreover, by exploiting the experience of similar surveys conducted previously. A scheme based on the practice of Finnish household surveys consists of the following six successive phases: (1) feedback from previous

surveys, (2) sample selection and updating, (2) interviewing and other data collection, (4) general analysis of nonresponse, (5) specific analysis of nonresponse, (6) comparisons of adjustments methods with other surveys (Laaksonen 1992, 18).

The most important features of each phase focusing on the needs of the estimation phase can also be listed. This means that the operations which will create the ideal conditions for the estimation are crucial. This leads, among other things, to gathering, updating and maintaining as much and as good auxiliary variables as possible, while the response rate should be kept as high as possible. The last point also means that much attention should be paid to attempts to provide information about problematic groups of respondents, including groups with special interest in this research. The basic concept is ignorability, originally presented by Rubin (see e.g. Little and Rubin 1987). A response mechanism is never ignorable over the whole data set, but we can go on toward sub-data sets where this principle may hold more or less true. These types of sub-data sets are called homogeneity groups or adjustments cells.

The improvement of ignorability calls for adjustments for nonresponse that Laaksonen has included in phases (4) and (5). Phase (4) concerns adjustments for unit nonresponse and phase (5) those for item nonresponse. For phase (4), his strategy is first to construct the so-called initial or basic sampling weights where the response mechanism is assumed to be ignorable with each pre-stratum. Secondly, population-level adjustments, typically post-stratification, could be used (presuming ignorability within post-strata). The next step could be to utilise sample-level adjustments for which purpose the response probability modelling could be applied (see Ekholm and Laaksonen 1991; assumptions: ignorability within adjustment cells and a model should fit well). Laaksonen (1991) has followed this strategy for the Finnish panel-based Income Distribution Survey.

Later, the so-called calibration methods (Deville et al. 1993) have become more common while helpful software such as CALMAR have been developed. Correspondingly, the 'ideal' strategy for adjustments for nonresponse and coverage errors have been revised by adding a fourth step, the calibration step, into the system. This last-phase calibration in particular, aims at providing the certain estimates as close to the other known figures as possible. For example, it is natural that some crucial population estimates from a survey are consistent with those derived from standard population statistics. This type of benchmarking information can also be drawn from another survey.

Although we had an ideal strategy for post-survey adjustments (on this term, see Groves and Couper 1995) it would not be possible to follow it in all situations. In particular, a tight time schedule is giving rise to another strategy, usually a simpler one. The use of any advanced methodology itself

is not the main reason for avoiding the best strategy, but instead the difficulties in introducing the ideal auxiliary variables into use quickly enough. In Finland, we have the following types of auxiliary variables:

a register variables derived from a sampling frame which are immediately available for use in adjustments but the quality of which may be weakened by some problems in updating ,

b other partly badly updated register variables which have normally been used for an initial sample, that is, covering both the respondents and the nonrespondents; in principle, these variables are immediately usable but have a low quality,

c really updated register variables often obtained from the taxation authority covering, i.a. taxable income, taxable wages, taxable property,

d certain other auxiliary variables which can be provided with additional effort and resources; the trap variables of this research are an example of these,

e external benchmarking information picked up from other surveys or specific statistics; the usability of these data is varying,

f intra-survey auxiliary variables which can immediately be used. The basic intra-survey variables, always used, are the indicators which classify the sample units into respondents, nonrespondents, overcoverage units and so on. Instead, a lot of other intra-survey information is not used at all although collected; for example such as the reason for nonresponse, contacting, duration of interview, point of time of interview, and partial replies,

g sample survey on nonrespondents which have rarely been performed and utilised at Statistics Finland.

We had a unique opportunity to exploit some variables of the group (d) in addition to the standard Finnish auxiliary variables of groups (a), (b) and (c). Since our exercise was a real post-survey test, the normal operations including adjustments for nonresponse had already been made. Therefore, we were not able to fully follow any ideal strategy, such as those mentioned above, but we added one step based on logistic regression and exploited the full-scale of auxiliary variables. This being the case, the following two sampling weights were already constructed: (i) the initial/basic weights, and (ii) the calibrated sampling (re)weights. For the comparisons of this research we only constructed the three alternative estimates for the starting panel of 1994. We do not consider the details of the sampling design (it is analogous to that used by Laaksonen in 1991). The sampling weights are as follows:

(i) The basic sampling weights $w_k(b)$ were thus constructed on the basis of the 26 pre-strata which are cross-classifications of 2 regions and 13 socio-economic groups (incl. taxable income groups one year earlier).

(ii) The calibrated weights $w_k(c)$ were calculated using CALMAR software (Deville et al. 1993) so that the auxiliary variable was the Finnish population classified by 13 regions, 8 age groups and gender.

It is assumed that the calibration weights will help in reducing the bias of several estimates, but it is easy to see that the variables of this procedure are fairly general, these do not include many important aspects of this survey. Therefore, we added the third phase which exploits a number of variables of nonrespondents as presented in Section 3.

(iii) These reweights $w_k(m)$, called model based reweights, are constructed from the reweights $w_k(c)$ multiplying these by the weights g_d as follows: $g_d = b/p_d$ in which p_d is an estimated response probability for the respondent of type d, and b is a coefficient so that the mean value of the weights g_d is equal to 1. These reweights are an extension for the weights presented by Ekholm and Laaksonen (1991).

Correspondingly, the variance estimates of the case (iii) consist of the two components: an ordinary component such as used in the Horvitz-Thompson estimators, and a component due to responding. The last component is zero, if the response rate is 100 per cent. However, the third component derived from the error term of logistic regression should still be added. The better fitting a model is, the less considerable this component is. We do not discuss this question in detail here, but give some comparable findings in Table 2 of results on different areas of the SLC. The general tendency of comparisons shows that the changes in estimates, when the basic weights have been converted to the calibrated ones, are more significant than the further changes after the model-based weights.

The results in Table 2 are estimated by the three weighting coefficients. The classes of the variables have been chosen to indicate the position "worse off". Results show that the nearer the described variable is to the variables chosen to the advanced model, the more the results change. Examples of these kinds of variables are marital status, education and the region/household size variables (see Annex 1a).

The general outcome in Table 2 is, however, that differences between different weights are slight at least by men and by women. There is yet one exception. All variables describing incomes; e.g. personal wage income, entrepreneur income, taxable income, and also incomes of property by men decrease when calibrated weights are used, compared to the basic weights. And again the average incomes decrease on average when adjusted weights are used. On the other hand, equality between people will worsen if the estimates obtained via more advanced methods are used. This will be seen via coefficients of variation, for example.

But as regards the activity and health variables, differences in results are so small that they do not violate the interpretation of Finnish well- of ill-being.

Table 3. Some SLC- estimates by men and women using the basic, calibrated and model based weights.

Variable/class	Men Basic weight	Calibrated weight	Model based weight	Women Basic weight	Calibrated weight	Model based weight
1) Variables used for adjusting						
Marital status: single (%)	36.0	38.8*	39.6*	29.3*	31.7*	31.5*
Education: lowest level (%)	30.5	30.9	*34.1*	33.5	33.2	*36.0*
Reg/size: Helsinki & households.=1(%)	2.50	2.35	*3.03*	4.48	4.33	*5.17*
2) Economy related variables						
Economical activity:	38.1	36.4*	*38.2	39.5	35.7*	*37.5*
wage-earner						
agricultural entr.	4.59	4.30	*2.00*	2.14	3.90*	1.90*
entrepreneur oth	5.68	5.40	*2.50*	2.54	5.40*	2.60*
Taxable income (mean, FIM)	92800	88700*	*87400*	68400	66300*	*65900*
c.v. (%)	82.3	87.5*	*86.9*	74.0	78.0*	*76.9*
Times of being unemployed during 5 years (mean)	2.38	2.37	2.37	2.02	2.01	2.00
Living width of residence: poor (%)	18.6	19.2	19.1	14.8	14.7	14.5
Has been at least one day without food because of no money: yes (%)	2.20	2.10	2.33	1.43	1.38	1.43
3) Health						
Evaluation of own health: poor (%)	7.56	7.25	7.55	7.46	7.22	7.55
Feels stress: continuously (%)	8.34	7.94*	8.03*	9.26	9.00*	9.09
Victim of violence (mean of number)	0.227	0.236	0.238	0.287	0.289	0.282
Antonowsky coherence score (mean)	67.1	66.9	66.8	66.5	66.3	66.3
4) Social relations						
Feels him/herself lonely: yes (%)	4.87	4.65*	4.86	6.96	6.77*	6.90
Number of friends (mean)	5.15	5.17	5.18	3.78	3.80	3.81
Could get a 200 US$ loan from a friend: no (%)	14.3	15.3*	15.8*	14.2	14.4	14.9
5) Activity						
Participation in associations: no (%)	61.5	61.1	62.2	66.3	65.9	66.8
Positions of trust: no (%)	72.3	72.5	73.5	78.2	77.3	77.9
Participating sports and recreational activity: no (%)	12.5	12.2	12.8	11.7	11.7	12.2
Participating culture and light entertainment: no (%)	28.6	27.7	28.5	21.3	20.7	21.2

* after the number = the calibrated estimate differs significantly (95 %) from the basic weight estimate
* in front of the number = the model based estimate differs significantly (95 %) from the calibrated estimate.

5. Conclusion

This research clearly speaks for the use of all auxiliary data that are available for post-survey adjustments, as they in particular may improve structural estimates, such as the number of poor people or less educated. Furthermore, the phenomena arisen from the structural effects may be analysed more correctly if good 'hard' auxiliary variables have been exploited. Many 'softer' factors, such as opinions and feelings, are, however, more complicated, at least insofar as the effects of those are not observable via 'hard' auxiliary variables. If good adjustments on these features are of high value for users, this leads to the conclusion that the auxiliary variables via a post-survey on nonrespondents should be used unless this information can be gathered in close connection with a normal survey.

It seems obvious that the conclusion made by crosstabulation, namely that nonresponse is in such a way harmless that it does not strongly distort the results, also holds with the models. But nevertheless the comparison of the results between men and women do not exclude the possibility of finding substantial differences in other population groups.

References

Ahola, A., Djerf, K., Heiskanen, M. and Vikki, K. (1995). Elinolotutkimus 1994. Aineiston keruu. (Survey on Living Conditions 1994. Goals and Field Work) Muistio 1995:2. Tilastokeskus: Helsinki. (in Finnish).

Deville, J-C., Särndal, C-E. and Sautory, O. (1993). Generalized Raking Procedures in Survey Sampling. *Journal of the American Statistical Association* 88, 1013–1020.

Ekholm, A. and Laaksonen, S. (1991). Weighting via Response Modeling in the Finnish Household Budget Survey. *Journal of Official Statistics* 7, 325–337.

Groves, R. M. and Couper, M. P. (1995). Theoretical Motivation for Post-Survey Nonresponse Adjustment in Household Surveys. *Journal of Official Statistics* 11, 93–106.

Heikkilä, M. (1990). *Poverty and Deprivation in a Wellfare State. A study of poverty and the accumulation of wellfare deficits in Finland.* Publications of the National Board of Social Wellfare and the Central Statistical Office of Finland. Valtion painatuskeskus: Helsinki. (in Finnish, English summary).

Laaksonen, S. (1991). Adjustment of Non-Response in Two-Year Panel Data. *The Statistician* 40, 153–168.

Laaksonen, S. (1992). *Handling Household Survey Nonresponse.* The Finnish Statistical Society. Research Reports 13. Haka-paino: Helsinki.

Lahelma, E., Manderbacka, K., Rahkonen, O. and Sihvonen, A-P. (1993). *Ill-health and its social patterning in Finland, Norway and Sweden.* National Research and Development Centre for Wellfare and Health. Research Reports 27. Gummerus: Jyväskylä.

Little, R. and Rubin, D. (1987). *Statistical Analysis with Missing Data.* John Wiley & Sons.

Salmi, M. (1991). *Ansiotyö kotona – toiveuni vai painajainen. (Working for wages at home – a dream or a nightmare)* Helsingin yliopiston sosiologian laitoksen tutkimuksia No. 225. Yliopistopaino: Helsinki. (in Finnish).

Vogel, J. (1991). *Social Report for the Nordic Countries. Living Conditions and Inequality in the late 1980's*. Statistical Reports of the Nordic Countries. The Nordic Statistical Secretariat: Copenhagen.

Annex 1a: List of explanatory variables in logistic regressions:

GENDER (SEX): 1 = Male, 2 = Female,
MARITAL: 1 = Non-Married (Never Married), 2 = Married, 3 = Divorced and Widoved

Combined Variables from MARITAL and GENDER by cross-classifying: MARISEX1 (11), MARISEX2 (21), MARISEX3 (31), MARISEX4 (12), MARISEX5 (22), and MARISEX6 (32) as the reference group.

REGION: 1 = Helsinki, 2 = Other South, 3 = Middle, 4 = North
HHSIZE: 1 = One-person household, 0 = Others
Combined variables from REGION and HHSIZE by cross-classifying: REGSIZE1 (11), REGSIZE2 (21), REGSIZE3 (31), REGSIZE4 (41), REGSIZE5 (01), REGSIZE6 (02),
REGSIZE7 (03), and REGSIZE8 (04) as the reference group.

AGE = Age in Years

EDUC1: 1 = Low Education, 0 = Others
EDUC2: 1 = Medium Education, 0 = Others
(the reference group in models)
EDUC3: 1 = High Education, 0 = Others

WAGE1: 1 = First Quartile Wage and Younger than 29 years, 0 = Others
WAGE2: 1 = First Quartile Wage and 29 years at least, 0 = Others
WAGE3: 1 = Second and Third Quartile Wage, 0 = Others
WAGE4: 1 = Fourth Quartile Wage, 0 = Others
(the reference group)

SOCEC1A: 1 = Manual Workers, 0 = Others
SOCEC1B: 1 = Non-Manual Workers , 0 = Others
SOCEC2: 1 = Entrepreneurs, 0 = Others
SOCEC3: 1 = Farmers, 0 = Others
SOCEC4: 1 = Retired, 0 = Others

UNEMP: 1= Breadwinner of a Household has been Unemployed during 1992,
0 = No

POORINCO: 1 = Poor Income, 0 = Others

ESUPPORT: 1 = Received Living Allowance, 0 = No

Annex 1b: Results on the three logistic regression models

Explanatory Variable	Parameter Estimates (and their Standard Errors) Panel 1993 to 1994 First Wave 1993	Second Wave 1994	Cross-Section 1994
MARISEX1	.349 (.166)	.496 (.182)	.568 (.133)
MARISEX2	.328 (.171)	.067 (.196)	.304 (.136)
MARISEX3	.089 (.153)	-.035 (.169)	.284 (.122)
MARISEX4	.205 (.153)	-.218 (.173)	.273 (.123)
MARISEX5	.607 (.181)	.105 (.226)	.275 (.160)
REGSIZE1	.684 (.187)	.519 (.225)	.798 (.134)
REGSIZE2	.362 (.135)	.123 (.157)	.558 (.119)
REGSIZE3	-.324 (.196)	-.791 (.252)	-.531 (.268)
REGSIZE4	.060 (.251)	.152 (.274)	-.372 (.268)
REGSIZE5	.433 (.149)	.784 (.160)	.823 (.104)
REGSIZE6	.169 (.091)	.064 (.104)	.387 (.080)
REGSIZE7	-.318 (.102)	-.419 (.119)	-.459 (.092)
AGE	.010 (.004)	.005 (.004)	.004 (.003)
EDUC1	.315 (.087)	.305 (.100)	.511 (.070)
EDUC3	-.130 (.123)	-.250 (.144)	-.470 (.111)
WAGE1	-.198 (.218)	.092 (.223)	-.437 (.146)
WAGE2	.277 (.194)	.644 (.226)	.222 (.177)
WAGE3	-.028 (.094)	.083 (.110)	-.037 (.268)
SOCEC1A	.008 (.108)	-.201 (.126)	.079 (.090)
SOCEC1B	-.168 (.135)	.088 (.146)	.039 (.105)
SOCEC2	.077 (.129)	.095 (.150)	.272 (.118)
SOCEC3	-.279 (.152)	-.616 (.191)	-.230 (.141)
SOCEC4	-.360 (.168)	.050 (.188)	.050 (.140)
UNEMP	.111 (.096)	.080 (.110)	.026 (.079)
POORINCO	.362 (.134)	.110 (.164)	.165 (.113)
ESUPPORT	.173 (.128)	.303 (.145)	.198 (.109)
Concordant	61.2%	63.4%	64.8%
Gamma	.238	.280	.304

A REASONED ACTION EXPLANATION FOR SURVEY NONRESPONSE[1]

Joop Hox
Department of Education, University of Amsterdam/Interuniversity Graduate School of Psychometrics and Sociometrics (IOPS)
Edith de Leeuw[2]
Department of Social Research Methodology, Vrije Universiteit te Amsterdam/IOPS
Harrie Vorst
Department of Methodology/Psychology, University of Amsterdam

Survey nonresponse is a threat to the inferential value of the survey method. To adequately fight the nonresponse problem extensive knowledge about respondents and nonrespondents is necessary. In this article we describe Cialdini's method to collect information of both respondents and nonrespondents. We show that this method can be successfully used in the Netherlands and that Cialdini's results have cross cultural validity. The decision to respond or refuse to a request for survey participation can be adequately described by the theory of reasoned action, but its predictive value is small.

Key words: Theory of reasoned action, nonresponse, response, survey, experiment, Cialdini, cross-cultural replication.

1. Introduction

In fighting nonresponse, two strategies can be adopted. The first is to increase participation through improved fieldwork methods. The second is to use postsurvey adjustment to compensate for nonresponse. These strategies are not mutually exclusive and both can be applied in one survey. To use coping strategies for nonresponse effectively, we must know more about differences between nonrespondents and respondents, especially in

[1] This chapter is based on a more detailed report in Bulletin de Méthodologie Sociologique, see Hox, de Leeuw and Vorst, 1995.

[2] The second author has contributed in the context of the NESTOR research program 'Living arrangements and social networks of older adults', conducted by the Vrije Universiteit Amsterdam and the Netherlands Interdisciplinary Demographic Institute in The Hague.

their attitudes and beliefs about survey participation. Experimental studies in the U.S. by Cialdini, Braver and Wolf (1991, 1993) and Cialdini, Braver, Wolf and Pitts (1992) showed that attitude toward surveys, perceived survey participation of friends, and value of privacy are consistently important factors in survey participation. A correlational study by Mathiowetz (1992) on U.S. Census data replicated these findings.

One general model to predict behavior is the theory of reasoned action (Ajzen and Fishbein 1980). The theory of reasoned action hypothesizes that the intention to perform a specific behavior is the only direct predictor of that behavior. This behavioral intention is an accurate predictor of behavior only if three conditions hold: 1) the intention and behavior measures correspond closely, 2) the intention does not change in the interval between measurement of intention and the occurrence of the behavior, and 3) the behavior is under the individual's personal control.

The intention to perform a specific behavior is in turn determined by the individual's attitude toward that behavior and the subjective norm. The attitude is formed by relevant beliefs about the consequences of the behavior, and the subjective norm is formed by the subjective perception of what relevant others think the individual should do.

The purpose of this study is a) to investigate whether the Cialdini findings replicate cross-culturally, b) to offer an integration of the findings using the theory of reasoned action, and c) to test whether our data fit a model of survey participation that is derived from the theory of reasoned action.

In the method section we first describe the paradigm for nonresponse research as developed by Cialdini, next we describe the actual data collection in the Netherlands. We follow this up by a short description of the measures used to operationalize the constructs in our application of the theory of reasoned action, and by a description of the analysis method used. In the result section we present an extended model that adds survey specific factors to the general theory of reasoned action model. We end with a discussion of the findings.

2. Method

Research Paradigm

To understand the causes of nonresponse we need information on both responders and nonresponders, especially concerning their attitudes and beliefs about survey participation. However, acquiring information on these factors poses a fundamental problem: nonresponders are people who do not want to provide information to researchers or are not available to provide information. Attempts to collect data from survey nonresponders is costly and often unsuccessful; the data collected from the small group of converted nonrespondents may not be representative for the whole total group of

nonrespondents, and the reasons given for not responding are retrospective and may be inaccurate through faulty memories or socially desirable answers (Cialdini et al. 1991).

To overcome this problem and probe attitudes and beliefs of both survey responders and survey nonresponders Cialdini developed a new experimental paradigm. This approach has three steps: 1) collect data from a 'captive audience' (i.e., a sample that will give 100 % response) on attitudes, beliefs, and personality characteristics, which might be related to survey participation; 2) implement a survey with this captive audience as sample, and register who is a respondent and who a nonrespondent; 3) try to model the response behavior on the survey in step two using as explanatory variables the data (available for both respondents and nonrespondents) from step one.

Cialdini's captive audience were American students in introductory psychology, who were required to fill in the response questionnaire as part of their first year program. Our captive audience is similar; in October 1992 we collected data from the freshmen at the department of psychology of the University of Amsterdam, who are obliged to participate in experiments and fill in questionnaires and tests as part of their program.

Data Collection

Step 1: Data collection with a captive audience. We collected data from the freshmen at the department of psychology of the University of Amsterdam. All freshmen students were presented with a description of a mail survey (vignette) and required to complete five questions on their response intentions regarding this hypothetical mail survey. In fact this vignette described in general terms the mail survey that would be administered in step two. The first question following the vignette concerned the direct response intention (i.e., if you receive the questionnaire and the request described here what would you do). The next five questions concerned the response intention in special situations (e.g., you do not have time at this precise moment what would you do). Two indicators were constructed based on these data: response intention (question one) and generalized response intention (questions two to six). People who score high on the index 'response intention' intend to return the hypothetical questionnaire; people who score high on 'generalized response intention' generally intend to return the questionnaire even if they have no time, important other obligations etc. Besides questions on response intention additional questions were asked regarding the attitudes and beliefs toward the hypothetical survey described in the vignette. Three weeks later, all students completed a selfadministered questionnaire, based on the questionnaire developed by Cialdini et al. (1991, 1992). This questionnaire contained questions on attitudes to survey research and on personal norms about response behavior and on the norms of friends.

Step 2: Survey implementation. In May 1993, six months later, a mail survey was sent to all freshmen in psychology of the University of Amsterdam. The topic of the questionnaire was study success and career expectations, the same topic that was described in the vignette in step one. In this questionnaire questions were asked about well-being, opinions on study, study habits, effective studying, grades, expectations about the future and about success of study and professional career. For this mail survey an adapted TDM approach (Dillman 1978; De Leeuw and Hox 1988) was used that omitted the last mailing by certified mail.

Step 3: Matching data and analysis. Very strict procedures were followed to match the data from step one to the response data of step two. For privacy reasons, the matching was done by a third party. This resulted in a sample of 462 cases.

Instruments

The *response behavior* was measured with a dichotomous variable that indicated whether one had completed and returned the mail questionnaire. Based on the self administered questionnaire from step one several multiple item indicators were constructed for behavioral intention, attitudes, norms, and beliefs. Also situational specific indicators on attitudes, norms and beliefs were constructed based on the vignette questions in step one. These indices are discussed in more detail below.

Two indices for *behavioral intention* were constructed based on the responses to the vignette: individuals who score high on 'generalized response intention' intend to return the questionnaire even if they have no time, important other obligations etc. Individuals who score high on the index 'specific response intention' intend to return the hypothetical questionnaire.

Attitudes: General attitude toward survey research was measured with eight questionnaire questions on the affective feeling toward survey research. Situational specific attitude toward this specific survey was measured with five vignette questions (e.g, how enjoyable would it be completing the specific questionnaire described).

Subjective norms: Internal (own) norms on survey participation were measured with four questionnaire questions. In these questions one was asked how probable it was that one would respond to general surveys from the government, universities, or marketing research societies (probability of responding).

Normative beliefs: Normative beliefs about survey participation were measured with three questionnaire questions on perceived friends' norms, asking estimates of the percentages of friends who would respond to different types of surveys.

Behavioral beliefs: Three multi-item scales were constructed that measured general behavioral beliefs about survey participation. The first involved the personal value of privacy. The second scale, positive experience, consisted of four questionnaire questions about the last survey experience. The third scale involved beliefs about the amount of effort needed for participation.

Two one item-scales were used to measure specific survey behavioral beliefs; they were based on questions asked about the vignette that described in hypothetical terms the survey that actually took place half a year later. These two scales were: Privacy beliefs specific for the survey undertaken (How large do you estimate the probability that in the described survey your answers will be treated confidentially) and beliefs about effort necessary for the survey (Can you spare the effort to complete and return the described questionnaire?).

Statistical Model and Analysis Procedure

Our model derives from the Ajzen-Fishbein theory of reasoned action (Ajzen and Fishbein 1980). The central constructs in the theory of reasoned action are the behavior, the behavioral intention, the attitude toward the behavior, the beliefs forming that attitude, the subjective norm, and the normative beliefs forming that subjective norm, specifically the perceived norms of relevant others. For personal behaviors the attitude will generally be the dominant predictor, and for social behaviors the norm will generally be the dominant predictor. We use the theory of reasoned action to specify a causal path model to predict response behavior. The basic structure of this model follows the theory of reasoned action, with behavioral beliefs added as predictors of attitudes, and friends norms as predictor of the personal subjective norm. To further explore the role of the attitude in this application of the theory of reasoned action, we specified the attitude part of the model twice, once for attitude toward surveys in general, and once for the attitude toward a survey such as the one actually used to measure the response (specific attitude). The available response indicators (intention to respond and response) are used as indicators of a latent factor 'intention to respond.' In our extension of the theory of reasoned action model, we use *two* indicators for attitude: one indicator for the attitude toward surveys in general, and one indicator for the attitude toward the specific survey actually used to measure the response. For a graphical representation of the model see the path diagram in the 'results' section.

Our model is a structural equation model with one latent variable 'response intention.' The scale of this latent variable is identified by fixing the path coefficient for general intention to one. The usual Maximum Likelihood estimators for structural equation models assume continuous data and multivariate normality. In our case, the dependent variable is a dichotomy, one predictor is a trichotomy, and two predictors have only five categories. Jöreskog and Sörbom (1989) recommend polychoric correlations and weighted least

squares estimation for such data. However, polychoric correlations assume that the categorical variables reflect underlying variables with a normal distribution. Preliminary analyses showed that this is not a valid assumption for our data. Instead, we use asymptotically distribution free (ADF, cf. Browne 1984) estimators, which do not make assumptions about the distribution of the data. Since ADF estimation requires large samples, we check the asymptotic standard errors and significance tests with bootstrap methods. A total of 1000 bootstrap samples is used to estimate bias-corrected standard errors (Stine 1989), and Bollen and Stine's (1992) adjusted bootstrap method is used to estimate the goodness-of-fit of the model.

3. Results

The basic model has a modest fit (chi-square is 100 with 30 degrees of freedom, p = .00). The modification indices for this model suggest two additions to the model. The first is a direct path from effort to subjective norm, indicating that a large perceived effort leads to lower assessment of the normative requirement to participate. The other is a correlated error term for both attitude measures. Both model modifications are theoretically reasonable. The resulting model fits well. The chi-square is 35.7 with 28 degrees of freedom. The asymptotic p-value for this chi-square is p = .15; Bollen and Stine's adjusted bootstrap estimates the p-value as p = .29. The goodness of fit indices are also acceptable: both GFI and AGFI are equal to 1.00, and the more strict Bentler and Bonett index is .93. On the basis of the significance tests and the goodness-of-fit indices we accept this model.

Table 1. Unstandardized path coefficients and p-values.

Path			Coefficient	p (asympt)	p (bootstrap)
attitude	<--	privacy	−0.13	.00	.00
attitude	<--	effort	−0.13	.00	.00
own_norm	<--	frnds_norm	0.73	.00	.01
attitude	<--	pos_exp	0.14	.00	.01
attit_spec	<--	priv_spec	0.42	.00	.01
own_norm	<--	effort	−0.50	.00	.00
intention	<--	own_norm	0.05	.00	.00
intention	<--	attitude	0.02	.26	.34
intention	<--	attit_spec	0.04	.00	.00
gen_intent	<--	intention	1.00 (fixed)		
spec_intent	<--	intention	2.92	.00	.00
response	<--	intention	0.37	.00	.00
attitude	<--	own_norm	0.08	.00	.00
own_norm	<--	attitude	0.56	.00	.00
atti_spec	<--	own_norm	0.32	.00	.00
own_norm	<--	attit_spec	0.05	.03	.05

Figure 1. Final model for survey response, standardized coefficients.

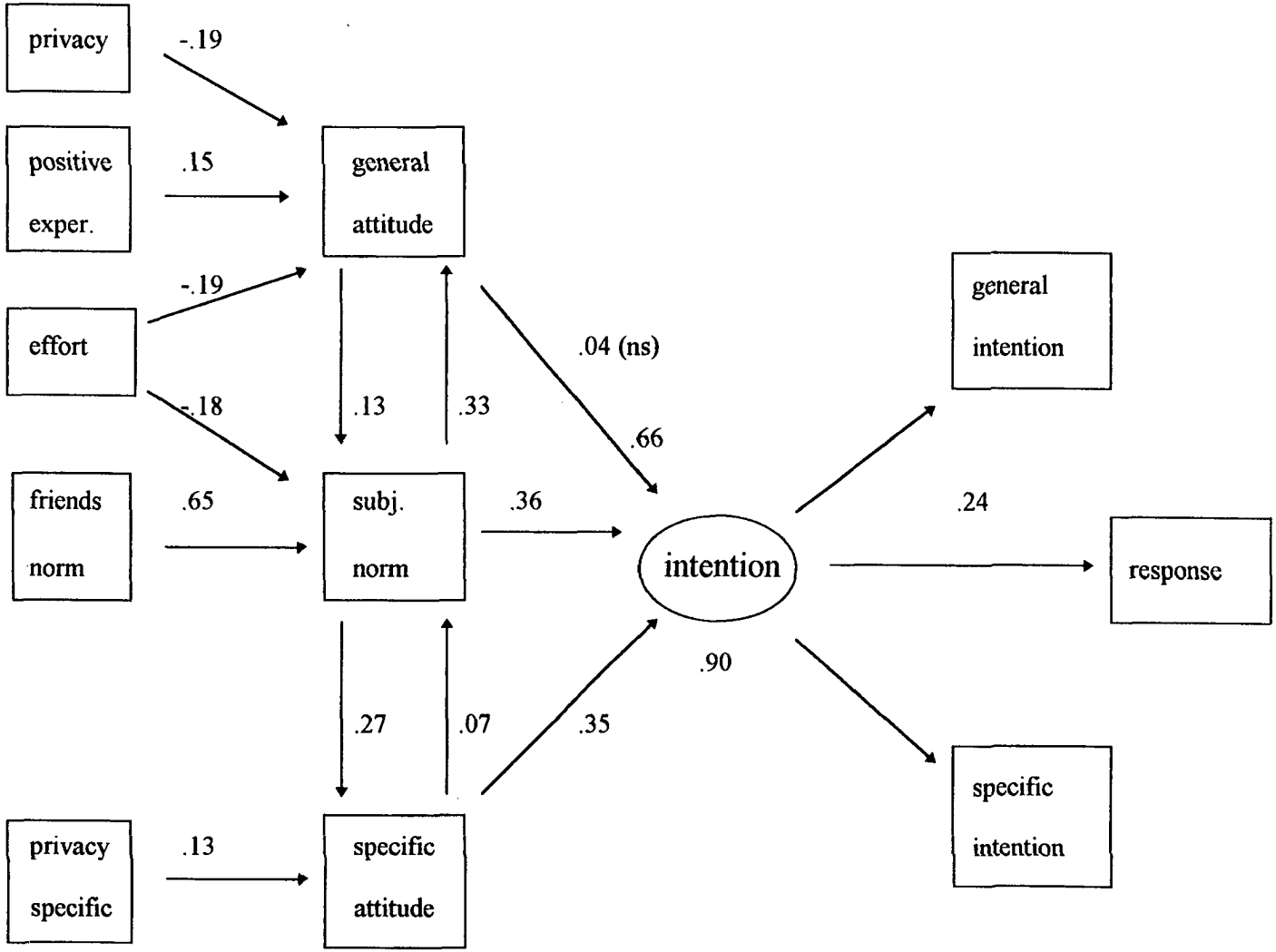


Table 1 presents the parameter estimates for this final model, with two sets of (one-sided) p-values: one set based on asymptotic standard errors, and the other on bias-corrected bootstrap standard errors. Most path coefficients in Table 1 are as predicted by the theory of reasoned action. The bootstrap standard errors tend to be larger than the asymptotic results, but all significance tests point to the same conclusion. The one path coefficient that is clearly nonsignificant is an extremely interesting result, because it is the path coefficient of general attitude toward surveys to intention to respond. The specific attitude toward the survey has a significant path to intention, but the general attitude not. Apparently, the intention to respond to a survey request depends more on the attitude toward that specific survey, and less on a global attitude toward surveys in general.

Figure 1 shows the graphical representation of the path model, with values of the standardized coefficients.

The standardized path coefficients in Figure 1 make clear that, although the model as a whole fits well, it does not predict the response very accurately. The path coefficient from intention to response is only .24, which is a weak relationship. The problem is not in the model itself; the ordinary multiple correlation of all predictors with the dependent variable is also a low R = .25. The requirement, that intention to respond is the only predictor of the response, does not reduce the correlation by a significant degree.

The subjective norm and the specific attitude are in terms of the standardized path coefficients equally important in predicting the intention to respond. It is interesting to note that in the reciprocal relationship between attitude and subjective norm, the subjective norm influences the attitude more than the attitude influences the subjective norm. Given the additional indirect path from the subjective norm through attitude to the intention to respond, the subjective norm is in fact a little more important than the attitude.

4. Discussion

The final model (Figure 1) fits well. However, it does not perform so well in predicting the final criterion, the response behavior. As we stated above, the theory of reasoned action is assumed to hold only if the intention and behavior measures correspond closely, the intention does not change in the interval between measurement of intention and the occurrence of the behavior, and the behavior is under the individual's personal control. Since one of the measures for behavioral intention is based on a vignette describing precisely the survey request actually used, we assume that the correspondence between the intention and behavioral measures is sufficient for our goal. Also, although problems such as time pressure at the time of the survey request may exist, we assume that the response behavior is

essentially under volitional control. However, since there is a time lag of six months between the first assessment and the final survey request, there is a considerable possibility that the behavioral intention has changed in that time period. This could be caused by a drop in saliency of the subject matter (study success and career expectations); the saliency can change strongly in the first year of study. In the Netherlands, Psychology is a popular 'temporary' study for students who have been barred from the study of their first choice, because of restriction of student numbers in certain studies (e.g., medicine). At the end of their first year in psychology they try again to enter the preferred study and they are no longer interested in their study success in psychology. Also, psychology is a favorite study for students who have not decided yet about their future, but have a general feeling that psychology is interesting and useful for their personal growth. These students often drop out during the year, but stay registered as students because they want to enjoy the financial and other student facilities. For such students, the salience of the questionnaire is low at the time the real survey was sent. Unfortunately, we have no way to detect these students, and incorporate them as a separate group in the analysis.

In our model, the correlation between the latent variable 'behavioral intention' and the final response is 0.24. In Cialdini's first two studies (Cialdini et al. 1992) seven variables significantly predicted response; no beta weights or correlations with response are given. In his third study (Cialdini et al. 1993) effects sizes are mentioned. No correlations are higher than 0.20, with a correlation of 0.16 between an indicator of response intention and the actual response. Clearly, the predictive validity of our theoretical model is not appreciably lower than the empirical regression model Cialdini uses.

In an explorative procedure, it is interesting to investigate if there are other variables than can predict reponse in our experiment. From the tests and questionnaires that our freshmen psychology students had to fill in, we selected 14 personality and mood tests, and 7 cognitive tests. The 14 personality tests comprised 111 sub-scales. Only eight of the 111 scales correlated at a p-level of (less than) .05 with the observed response. If we apply a Bonferroni correction, none of these correlations are significant at an alpha of 5 percent. Thus, contrary to popular belief, there appears to be no such thing as a 'responding personality.' Of the seven cognitive tests, three correlated at a p-level of (less than) .05 with the observed response. If we again apply a Bonferroni correction, two of these correlations (with 'verbal analogies' and 'drawing conclusions') are significant at an alpha of 5 percent. This outcome confirms the effect found by Bros et al. (this volume), that intelligence is a predictor of response in surveys. However, the effect is again small; the multiple correlation of the two cognitive tests with response is only 0.15, which is smaller than the 0.24 reached with our theory of reasoned action model.

References

Ajzen, I. and Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Bollen, K. A. and Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods & Research*, 21, 205–229.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62–83.

Cialdini, R. B. (1992). Personal communication. *The Hague, Third International Workshop on Household Survey Nonresponse.*

Cialdini, R. B., Braver, S. L., & Wolf, W. S. (1991, October). A new paradigm for experiments on the causes of survey nonresponse. *Paper presented at the Second International Workshop on Household Survey Nonresponse. Washington, DC.*

Cialdini, R. B., Braver, S. L., Wolf, W. S., & Pitts, S. (1992, September). Who says no to legitimate survey requests? Evidence from a new method for studying the causes of survey nonresponse. *Paper presented at the Third International Workshop on Household Survey Nonresponse. The Hague.*

Cialdini, R. B., Braver, S. L., & Wolf, W. S. (1993, September). Predictors of nonresponse in government and commercial surveys. *Paper presented at the Fourth International Workshop on Household Survey Nonresponse. Bath.*

De Leeuw, E. D., & Hox, J. J. (1988). Response stimulating factors in mail surveys. *Journal of Official Statistics*, 4, 241–249.

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method (TDM)*. New York: Wiley.

Hox, J. J., de Leeuw, E. D. & Vorst, H. (1995). Survey participation as reasoned action; a behavioral paradigm for survey nonresponse? *Bulletin de Méthodologie Sociologique*, 47, 52–67.

Jöreskog, K. G. and Sörbom, D. (1989). *Lisrel7. A guide to the program and applications*. Chicago: SPSS Inc.

Mathiowetz, N. A. (1992, May). A behavioral paradigm for understanding nonresponse to the 1990 Census. *Paper presented at the annual meeting of the American Association for Public Opinion Research. St. Petersburg, Fl.*

Stine, R. A. (1989). *An introduction to bootstrap methods: examples and ideas*. Newbury Park, CA: Sage.

TOWARDS A CURRENT BEST METHODS FOR NONRESPONSE RATE REDUCTION AT STATISTICS SWEDEN

Lilli Japec and Lars Lyberg
Statistics Sweden, Sweden

The purpose of this paper is to describe some initial findings from a survey conducted among 50+ surveys that regularly appear in Statistics Sweden's nonresponse barometer. The survey has been conducted to shed some light on procedures used in these surveys to reduce nonresponse rates. The survey is one of the first steps in developing what is called a set of Current Best Methods (CBM) for dealing with this problem in a standardised fashion across our agency. In order to provide some background we describe the nonresponse barometer and some earlier attempts at dealing with this problem. The CBM project work is described in general terms and after that findings on process components in the 50+ surveys like advance letters, data collection strategies, questionnaire design, interviewer training, respondent burden, and incentives are described. Initially it was assumed that these descriptions could lead to some kind of Pareto analysis where some crucial process steps were identified and emphasised in the CBM. It seems, however, as if all components need to be dealt with extensively.

Key words: Nonresponse barometer, TQM, internal nonresponse survey.

1. Current Nonresponse Trends at Statistics Sweden

The Nonresponse Barometer

Since 1985 Statistics Sweden has produced what is called the nonresponse barometer. The barometer simply describes nonresponse rates in continuing surveys across the agency. The surveys concern both surveys of individuals and households and surveys of businesses and establishments. The purpose of the barometer is that it should serve as an indicator of the survey climate, i.e., the willingness to participate in our surveys. Over the years the barometer has come to include more and more of the continuing surveys. Back in 1985 the barometer included only a few surveys of individuals and households. In 1995 it includes well over 50 surveys covering subject matter areas like labour market, education, agriculture, industry, investments, welfare, income, and salaries to mention just a few.

The nonresponse trends differ across subject matter areas. In some surveys, like the Labour Force Survey, the trend is that the rate is more or less continuously increasing, while other surveys show irregular patterns, with ups and downs. Some surveys even have nonresponse rate trends that are decreasing. Overall, however, most survey managers are not pleased with the current nonresponse rate levels. There is a need for improvement.

Some Examples

Several of Statistics Sweden's surveys covering the business and the establishment sector are mandatory. The repercussions for nonrespondents in those surveys are moderate fines but this route is seldom enforced by the agency since it can create considerable bad-will that can affect participation in other surveys. Nonresponse rates in mandatory surveys range from 2 to about 20 percent. Clearly this is not acceptable for surveys that government has classified as being particularly important.

The Labour Force Survey nonresponse trend is very clear. In the early 70's the rate was very low around 2 or 3 percent. Today it is approximately 13 percent and most of it consists of persons the interviewers have not been able to contact. Other surveys on individuals show higher rates around 20 percent, either because they contain sensitive questions like the Party Preference Survey or because they are extensive like the Survey of Living Conditions. The most troublesome survey is the Family Expenditure Survey with nonresponse rates steadily above 30 percent. Here the explanation is that respondent burden is considerable.

Some Conclusions from Continuing Measurements

The barometer is a valuable tool for top management, methodologists, and individual survey managers. One might argue that too much emphasis is put on measuring nonresponse. After all, there are many other specific error sources in a survey and the total survey error is what we strive to control. Focusing on nonresponse too much might contribute to a lesser interest in these other sources. Admittedly, this might be the case but the obvious solution is to start measuring other data quality indicators as well. There are some good reasons for looking hard at nonresponse rates, though. First, nonresponse rate is such a visible and easily understood indicator of data quality. Most survey users and sponsors know that large nonresponse rates might have an impact on estimates. Working with this error source should be an efficient way of teaching data quality to survey data sponsors and users. Second, at Statistics Sweden we feel that nonresponse rates reflect the survey climate. By keeping track of the nonresponse rates we know when respondent relations deteriorate and we can notice the impact specific events have on the survey climate. For instance, the very project we are describing

in this paper is a result of this procedure. We feel that the nonresponse rate is one of the most important factors that our agency is judged by.

It is evident that specific events can have a tremendous effect on nonresponse rates in several surveys. For instance, in 1970 confidentiality became an issue during the census data collection. This single event was the starting point for generally increasing nonresponse rates in Sweden. In 1986 there was a debate on the Metropolit survey, a survey that was not even conducted by Statistics Sweden. It was discovered that data were collected from registers for a sample of persons born on the 15th of each month without their consent. Immediately nonresponse rates in Statistics Sweden's surveys started climbing. Other debates concerning census-taking and the use of registers have also affected nonresponse rates. The general pattern is that after a while nonresponse rates have decreased but to a level higher than before these specific events occurred.

It is typical for Sweden that nonresponse rates increase during vacation periods, especially during the summer. In interview surveys it is also quite common that average nonresponse rates per interviewer vary a lot which is a fact that reflects some kind of skill effect. Some interviewers have astonishingly low nonresponse rates even though the survey might be difficult to conduct or that the areas they work in are difficult.

Thus there is a general pattern of increasing nonresponse rates and/or high nonresponse levels. Under this pattern there is another showing lots of variation between surveys, modes, geographic areas and between and within interviewers. It seems as if there is need for standardisation when approaching the problem of reducing nonresponse rates.

2. Examples of Earlier Work

Thus, nonresponse rates are a concern at our agency and they are constantly on the agenda. Three times special projects have been started involving many parts of the organisation. All three times the situation has been such that top management has judged it necessary to do something on a global scale.

The UBIS project

The 1970 census problems had by 1976 caused all time high nonresponse rates for most surveys of individuals and households. An agency project was started and its goal was to put an end to this increase and after that hopefully turn the trend downwards. The project was led by top management and involved large parts of the staff working on surveys of individuals and households. The project consisted of three parts: issues on information to respondents, strategies for data collection, and adjustment techniques. The

first sub project developed a handbook on respondent relations describing efficient ways of approaching the respondents via advance letters, brochures, questionnaires, and instructions. The handbook also dealt with pretests of information material, how to cooperate with media, and how to approach special population groups like immigrants and the elderly. The data collection sub project concentrated on issues related to tracing sample members, the mail survey collection process, and interviewers. The project on adjustment techniques developed a handbook on weighting and imputation and associated variance formulas. Issues on nonresponse definitions, co-ordination of sample selection, subsampling of non-respondents, and effect on estimates were also studied.

The project went on for four years and needless to say lots of other methodological work had to wait during that period because of the massive emphasis on nonresponse. What happened was that the people involved in the project, but apparently not the entire organisation, learned a lot about nonresponse, its causes and effects. And indeed, nonresponse rates were stabilised during that period but it is difficult to say how much of this improvement that could be attributed to project work. Later experiences have shown that nonresponse rates depend on media interest and once the media lose interest the public "forgets" to resist. Thus a certain stabilisation might have occurred even without this gigantic effort. In any case several new procedures were introduced as parts of the production process at Statistics Sweden but not in a standardised fashion. Survey managers always had the option not to change things.

Metropolit

Nonresponse rates for surveys of individuals and households decreased until 1982 when the survey climate changed due to the debate about the upcoming 1985 census. In 1986 when nonresponse rates were already increasing the Metropolit study was exposed by the media. This debate left the public so concerned about confidentiality and trust that nonresponse rates increased dramatically in most of Statistics Sweden's surveys.

It was absolutely necessary to do something about this situation. A task force was created and its mission was to propose measures to deal with the acute problem. The organisation of this work differs from UBIS. This task force consisted of 11 persons who worked during seven months. After that it was supposed that the various survey managers and the interviewing unit should implement the suggestions.

The task force suggested, among other things, increased experimentation on interviewer co-operation, reinterviews and interviewer training, studies of the effects of late respondents and noncontacts, revision of de-identification procedures and the reporting of nonresponse in publications.

We believe that this project model did not work very well. It was up to these other departments at the agency to operationalise and implement the suggestions provided by the task force. To our knowledge very little happened as a result of the project and this is not surprising given our experience with this kind of model. It is essential that those working on process improvements are those who actually developed the suggestions for these same improvements.

3. The current project

Statistics Sweden's TQM effort

Since 1993 Statistics Sweden has worked with quality issues according to the approach usually labelled TQM (Total Quality Management). The styles of TQM vary a lot between organisations that use this way of working but there are a number of common features like customer focus, process orientation, decisions based on data and everybody's involvement. The reasons why our agency had to start systematic quality work were many. First, the statistical system in Sweden has changed to become more decentralised. Over 20 agencies have received about half of the funds that Statistics Sweden once got and these agencies can now choose to let Statistics Sweden do the work, to let some other firm do it or they can do it themselves. So almost overnight Statistics Sweden has become a statistical firm of considerable size. We have to compete which is a new situation. Another reason is that funds that we still get will no doubt decrease over time so we have to do more with less funds in the future. A third reason is that the quality of our work actually has to increase. Survey users will become more aware of quality issues in the future and we must be prepared for that situation.

The TQM work at Statistics Sweden includes improvement project work. There are about 50 such projects going on and each project uses tools like flow charts, Pareto diagrams, control charts and fishbone diagrams. TQM projects are based on team work and each team has a facilitator. In the long run it is supposed that all project work at Statistics Sweden should be run like these projects. Another line of work is the development of 23 overall strategies that should govern our work. Notably, one of these strategies is that we should use current best methods in our work. A third line of work is a general evaluation of work at Statistics Sweden using the criteria that have been developed by the Swedish Quality Institute.

The Need for Standardisation

Two of the TQM projects deal with the development of CBM (Current Best Methods). These concern the areas of editing procedures and reduction of nonresponse rates. Both of these are examples of large processes that are used by many surveys within the organisation. A general observation is that these processes are costly and performed in a nonstandardised fashion. It would be most beneficial for the agency if these processes could be made more standardised.

Goal of the Project

In 1994 top management decided that nonresponse rates at Statistics Sweden had reached an unacceptable level. The R&D Department was asked to do something about it and the department defined a TQM project with the mission to develop a set of best methods to reduce nonresponse rates, assist in implementing these methods and check whether any improvements occurred. Once again the solution seems to be to develop a handbook but this time there is also need for an implementation process that is standardised. Based on the initial findings from the internal nonresponse study (see Section 4) we have decided to keep the handbook relatively simple for reasons that are given in that section.

Project Organisation and Initial Experiences

The project organisation was initially very similar to the one used by the Metropolit task force. Representatives from various parts of the organisation were asked to join the project. An overview of reduction methods was given to the project members, a mission statement was agreed upon and certain areas were identified where work was needed. Some of these areas were definitions of rates, content and style of advance letters, and reasons for nonresponse. Work started in these areas. Pretty soon we discovered some problems associated with project member expectations. This caused us to form an inner group. The inner group consists of just six members who are supposed to gather material for the handbook, actually write it and use the rest of the group as a resource for special studies and reviews of handbook chapters. The group is supposed to deliver the handbook by April 1996 and this time special attention will be given to the promotion of the findings.

4. An Internal Nonresponse Survey

One of the first things the inner group did was to approach the final customer of the intended handbook. We really needed to find out what was being done at various surveys and evaluate the current state of affairs. It was

decided that the managers of the 50+ surveys that were included in the nonresponse barometer should fill out a questionnaire regarding measures taken to reduce nonresponse rates. The response rate so far is above 90 %.

Description

In the survey we ask about, for instance, length of data collection period, the writing of advance letters, choice of mode and combination of modes, number of contact attempts, strategies for terminating the data collection, sensitive questions, respondent burden, interviewer training, the use of proxy respondents, questionnaire testing and studies of nonresponse.

An Overview of Findings

Given all the work that has been laid down on nonresponse reduction over the years it is surprising that so little of what is known about methods in this field has been picked up by the survey managers. There is a general lack of data on nonresponse and nonrespondents that could guide survey managers in their work. There is also variation between surveys in the sense that some have more ambitious processes than others. Very few have put in great efforts in developing questionnaires and advance letters. Studies focusing on the respondents' (suppliers') attitudes and concerns about the surveys are rare even though the agency could benefit a lot from collecting such information. Also it seems as if there is some confusion regarding what should be considered an acceptable nonresponse rate. In most surveys the rate that has been achieved when the allotted time for data collection is up is the one that is accepted. In some cases a certain nonresponse rate is negotiated between the sponsor and Statistics Sweden and in those cases there is a strong tendency that the resulting rate only marginally exceeds the negotiated value. Some surveys use selective nonresponse follow-up so that one concentrates on, say, large missing companies or strata where initial nonresponse is largest. But some are obviously selective in another sense, namely they tend to include the "easiest" nonrespondents.

There seems to be some interest in reducing respondent burden although some survey managers have problems deciding whether the burden in their surveys is extensive or not. Surveys that use the interview mode seem to rely heavily on the Survey Unit when it comes to training and nonresponse follow-up, when in fact co-operation between them would probably generate better results.

There seems to be a great need for standardisation, just as was assumed when the CBM project started. Many data collection processes need to be improved using more data as a basis for rational follow-up procedures and using more of the methodology available in this field.

5. Advance Letters

By law Statistics Sweden has to send out an advance letter to inform the potential respondents about their rights and the purpose of the survey. If the survey is not mandatory this should clearly be stated in the advance letter and if other information is going to be obtained about the respondent through administrative records, the respondent has to be informed about this.

When studying the advance letters we found that many of them are old, i.e., only minor revisions have been made during the last five years and as far as we know only one advance letter study has been carried out during the same period. Most of our household and individual surveys are voluntary but the majority of our business and establishment surveys are mandatory. There is a clear difference in the tone used in the advance letters to voluntary compared to mandatory surveys. The advance letters sent out for mandatory surveys are often quite impolite and demanding. The fact that the survey is mandatory is sometimes used as a threat in the advance letter. (We have not yet looked at the reminders.)

The layout of the advance letters varies but many have a very grey and dull appearance and the language is often very bureaucratic. On the other hand we found letters that were very informal. In some of the letters the grammar is incorrect. Our findings have generated a new project with the task to develop recommendations for how to improve our advance letters.

6. Data Collection

Many of the data collection processes seem to be old processes that, when in need for improvement, have been built on rather than redesigned. It is clear that in some surveys the time for sending out reminders and the number of reminders have been decided without any data collection on incoming response flows. In our survey we found that in similar mail surveys the timetable for reminders and the number of reminders vary a lot. We found the same result in surveys where telephone reminders are used. The length of the data collection period does not seem to have been very carefully thought through in a number of surveys. There are reasons to believe that there is a lot of "air" in many data collection procedures. For instance, in the survey of work injuries the data collection period was recently shortened from 5 months to 2.5 months mainly by standardising the data collection process, and picking the "low-hanging fruits".

In most of the surveys the main reason for ending the data collection is that the time for publication of the results is approaching. But in some surveys we found specific goals, such as a response rate of 80 %. How those

goals had been decided was unclear in some cases but in other they were the result of negotiations between the sponsor and the interviewer unit.

Some of the business and establishment surveys use selective nonresponse follow-up. In household surveys, however, selective nonresponse follow-up is not used. The most common follow-up procedure in interview surveys is to let the interviewers themselves decide who to phone or visit. A problem with this procedure is that there is a risk that interviewers will concentrate or automatically get the "easy" cases. Clearly, such a procedure might lead to problems, because some groups like immigrants are much more difficult to reach than others and it is possible to construct examples where it could add to the bias. Total survey error should be in focus and therefore selective follow-up procedures should be tried more systematically. In three surveys a sample of the nonrespondents is followed-up according to the Hansen-Hurwitz subsampling plan.

7. Questionnaire Design

The questionnaires are usually developed by the survey manager. Most surprisingly, a vast majority of the questionnaires have not been tested by our cognitive laboratory. There are a few questionnaires that have been tested during the last five years but most of them are from the same subject-matter department. Only one of the questionnaires sent to establishments has been tested. In establishment surveys the companies' own administrative records are needed to fill in the questionnaires so the type of data that could be provided is looked upon as the main problem rather than the way the questions are asked and the way the questionnaire appears to the respondents. However, the results of a study of reasons for not responding in the Survey of Parishes, shows that many of those who did not respond said that the questionnaire was difficult to fill in.

There seems to be a lack of knowledge among survey managers whether questions are sensitive or not. Special methods developed for sensitive questions, e.g., randomised response or item count, are not used in any survey.

8. Interviewer Training

The interviewer organisation consists mainly of field interviewers but there is also a small central group of telephone interviewers. The interviewers make their own decisions on when to make call attempts. They work weekdays except for Friday evenings. Many of the field interviewers have been working as interviewers for many years. They have accumulated

knowledge over the years and make there own "optimal call-scheduling algorithms" depending on the background information they have on the person they are trying to reach. However there is still a large variation among interviewers and there is a need for standardisation.

In many telephone surveys the noncontact rate is very high. Statistics Sweden has developed a new CATI-system called WIN-CATI. With the new system it will be possible to collect information on the number of call attempts that have been made and when they were made. This information could be used to develop optimal call-scheduling algorithms to be used by all the interviewers.

The interviewers are usually given an initial product specific training. There is little knowledge among the survey managers about the interview process and what type of training the interviewers get to be prepared for refusal conversion.

9. Respondent Burden

Respondent burden is difficult to measure but there are four components that can indicate whether the respondent burden is heavy or not:

- the time it takes to answer the questions.
- whether additional material is needed to answer the questions.
- how often the respondent participates in our surveys.
- if there are sensitive questions in the survey.

The notion of respondent burden varies a lot. Here are some examples:

- In establishment surveys the respondent almost always needs special material to fill in the questionnaire. This fact is usually not considered as a factor that increases the respondent burden.
- An interview that takes 65 minutes with sensitive questions about health, economy and victimisation was considered by the survey manager to have a light respondent burden.

The respondent burden is extremely heavy in some surveys e.g., in an establishment survey questions on salaries and time variables for employees are asked and it could take up to three days to fill in the questionnaire. Data are not allowed to be sent in on a disk. For many medium-sized establishments the respondent burden is therefore very heavy. Another example of respondents with heavy burden is owners of apartment buildings. They can get hundreds of questionnaires to fill in for their apartments. The same sample is used for two other surveys so they are participating in three surveys where similar questions are asked. They also need additional files to fill in the questionnaires.

The respondent burden is discussed in some programmes but few measures have been taken to decrease it. Examples of measures that have been taken include; using data already available in administrative records, allowing item nonresponse, using cut-off, and taking new samples. In the above example of owners of apartment buildings a new sample will be taken at the next survey occasion to decrease the respondent burden. The government is continuously reminding Statistics Sweden to take measures to decrease respondent burden. Obviously, more needs to be done.

10. Incentives

At Statistics Sweden incentives are not used in many surveys and when they are given they are typically pocket calculators or a lottery tickets. The main reasons for not using incentives is the fear of respondents getting used to receiving incentives for their survey participation and that incentives will increase survey costs. Another issue is that incentives would perhaps increase survey participation among certain categories such as students and people with low income. Nevertheless, incentives have proven useful as a means of stimulating survey participation. All measures to increase participation are costly and there is no reason to single out the use of incentives as a method that should not be used.

11. Towards a CBM

Experience shows that it is very difficult to disseminate knowledge of statistical methods to the various programs at Statistics Sweden. Earlier instructions on for instance, nonresponse analysis, cognitive laboratory testing of questionnaires, and variance estimation have not been implemented to the extent that we expected. Therefore we need to try a different approach this time to reach our goal, i.e., to standardise procedures for nonresponse reduction.

The normal procedure has been to write a paper and distribute it to a group of methodologists for further dissemination. However, this is not sufficient. Part of our launching program this time is to provide compulsory courses on different levels, e.g., top management, survey managers at each program, staff at our interviewer unit, and chief statisticians. Besides this we will have a support team consisting of people from the project group. The support team will provide help at the implementation stage.

The CBM will be a handbook for survey managers and it is therefore important that it is written in a way so that the methods are easy to understand and could be used by someone who is not a methodologist. By

including graphics, checklists, and flowcharts we believe that this goal can be accomplished. One of the purposes of the CBM is that it should inspire to make experiments and improvements. Each chapter is to be developed further and updated with new findings in a forthcoming CBM.

For instance, the contents of the chapter on sensitive questions could be something like this:

A What is a sensitive question?
   *theory and examples
B How do I construct a question?
C Where do I place sensitive questions?
D What effect does the mode have?
E What special methods are there?
   *randomised response
   *item count
   *confidentiality, protection techniques
F What could be done with the advance letter?
G How do I deal with item nonresponse?
H How frequent are sensitive questions at Statistics Sweden?
I References (not more than three main references)

Other chapters will treat, among other things, strategies, respondent burden, questionnaire development, interviewer organisation, interviewer training, and advance letters. The CBM should not exceed 80 pages and the first version will be published 1996.

TELEPHONE COVERAGE AND ACCESSIBILITY BY TELEPHONE IN FINLAND

Vesa Kuusela
Statistics Finland, Finland

Telephone coverage, i.e. the percentage of households having or having access to telephone, is usually evaluated by general purpose surveys, like household surveys. It is commonly supposed that the actual frame population in telephone surveys is defined by the achieved percentage of telephone households and, in addition, that all people having a telephone can be reached. Accessibility by telephone in Finland is studied in this article. The results show that there was a gab of nearly 10 % between the generally accepted telephone coverage and the percentage of people to whom a telephone number could be traced. The results lead to a question whether the telephone coverage obtained in the general purpose surveys is sufficient in estimating accessibility by telephone. One reason is that in general purpose surveys there is usually fairly large non-response. In addition, there are many socio-economic, cultural and lifestyle variables that regulate the subscription to telephone services, making it a complex procedure. Therefore, telephone coverage and accessibility by telephone, to be accurate, should be defined in a specific survey.

Key words: Telephone coverage, accessibility by telephone, estimation of telephone coverage.

1. Introduction

In Finland 94 % of households have a telephone according to the Household survey. In other sources the telephone coverage is estimated to be 94 % – 95 %. Also in the United States approximately 94 % of all persons live in households with a telephone according to Brick et al. (1995). Thus it should be quite straightforward to expect that about 94 % of a random sample of the population could be contacted by telephone. That does not seem to be the case, not in Finland at least.

Usually, for a survey conducted by Statistics Finland, the sample is drawn from the population register by simple random sampling of individuals. The register is maintained carefully and its coverage of the population is considered satisfactory. Both the under-coverage and the over-

coverage are insignificant. The register does not include telephone numbers, however. Therefore, the only way to obtain the numbers is to search them from telephone catalogues. A couple of times the task has been commissioned to a telephone company maintaining or having access to telephone catalogue databases.

First the telephone number of the sampled person is searched for by his or her name and address. If a telephone number for the target person is not found, the search is extended to the two eldest persons living in the same dwelling.

Surprisingly enough, this procedure has yielded telephone numbers for only about 85 % of the sample, including roughly 3 % secret numbers. That is, there is a gap of nearly 10 % between the telephone coverage obtained from the Household survey and the proportion of persons to whom a telephone number could actually be attached. The difference is relatively big, particularly because it is partly surplus to the 'normal' non-response in telephone surveys.

In other countries several studies have been published dealing with the nature of the non-telephone households (e.g. Groves 1989, Smith 1990, Brick et al. 1995, Keeter 1995). In Finland this specific subject has not been looked into. However, after the discrepancy was observed, a small study was carried out to shed some light on the situation in Finland and on the reasons for the gap.

2. Method

A simple random sample of 5 785 individuals from the population register was analysed. The sample was composed of two separate samples for two different surveys undertaken in 1993. That is, the sample was not drawn specifically for this study. Telephone numbers were traced by a company maintaining the telephone catalogues. The analysis had to be made in quite general terms as only a limited number of variables were available. For instance, such central features as income and marital status were not available. The purpose was to only find some general determinants for the difference.

3. Results

A telephone number could be attached to 85.2 % of the sample including 3.1 % secret numbers.

Numbers were found more frequently for women than for men (86.8 % vs. 83.5 %). On the other hand, women had slightly more secret numbers

(3.6 % vs. 2.7 %). Altogether, a publicly accessible telephone number could be found for 83.2 % of the women and for 80.8 % of the men.

A telephone number was traced more often for older persons than for the younger ones. Figure 1 illustrates how the percentage of accessible telephone numbers varies with respect to the age and gender of the sampled person. The most problematic subpopulation seems to be people between 25–29 years of age, especially men. Evidently, in this subpopulation many young persons have moved from their parents' homes to homes of their own. Many of them are probably still finishing their studies in colleges or universities or they have just graduated. Furthermore, it is noteworthy in figure 1 that men less than 55 years of age have telephones below the average (only few men have secret numbers).

Fig. 1. Accessible telephone numbers in Finland by sampled person's age and gender.

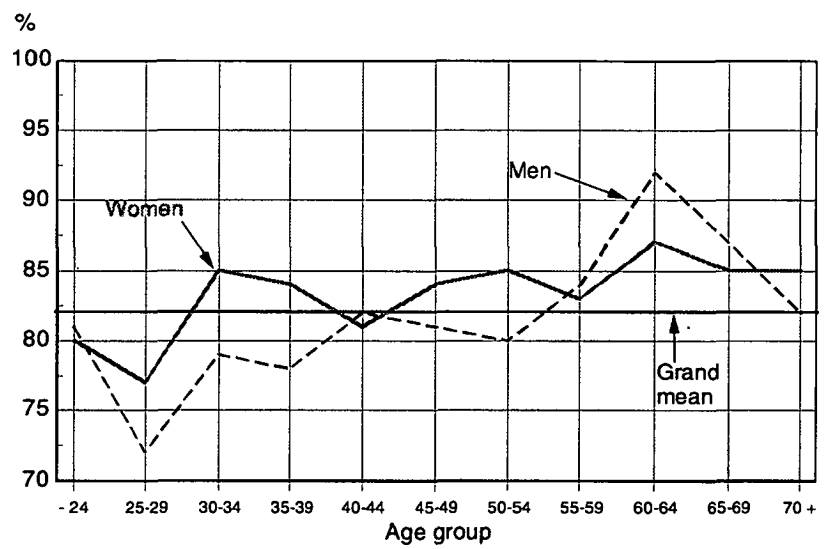
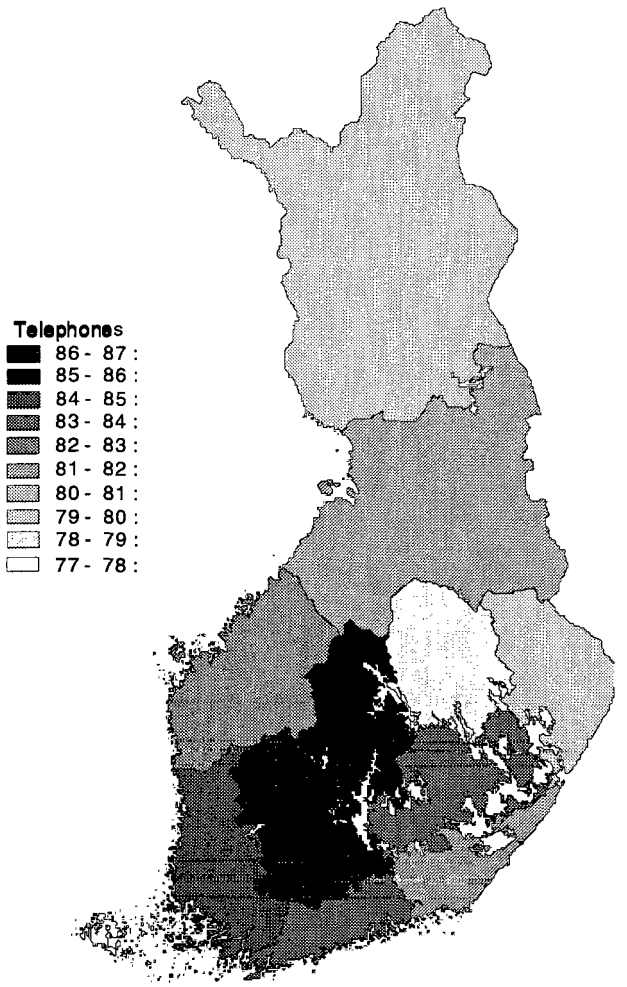


Figure 1 also reveals a slight interaction between age and gender: young men have much fewer telephones than young women whereas old men have more telephones than old women.

Geographically thinking, households with telephones are more frequent in the southern and south-western parts of the country on average (see figure 2). The difference is more than 8 % between the highest and lowest mean values: the highest province average was 86.2 % (in the province of Häme in the middle of the country) and the lowest average was 77.8 % (in the province of Kuopio a little Northeast of the province of Häme).

Fig. 2. The percentage of found telephone numbers in the provinces of Finland. Helsinki is excluded from the southern province.



The type of municipality is also connected to the proportion of non-telephone households. Overall, in towns the proportion of found telephone numbers was above average. However, there were some noteworthy exceptions (e.g. Helsinki, see figure 3). On the average, the percentage was highest in the small and middle sized towns. In the overall average the tendency was: the more rural-like the municipality was, the fewer telephone numbers were traced. Especially in eastern Finland the municipalities where agriculture is the most common means of earning a living there was a low telephone coverage. However, there was some variation in the province profile. For instance, one of highest telephone coverages (91.7 %) was in the most rural municipalities in Lapland (the most northern province).

Fig. 3. Proportion of households in major cities in Finland where there is a telephone, a secret number or no telephone.

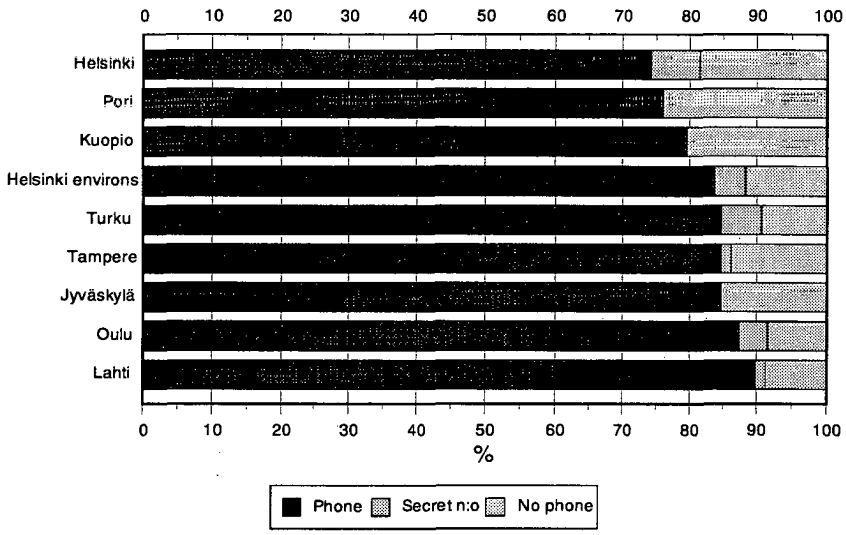


Fig. 4. Accessability by telephone by gender and age in Helsinki.

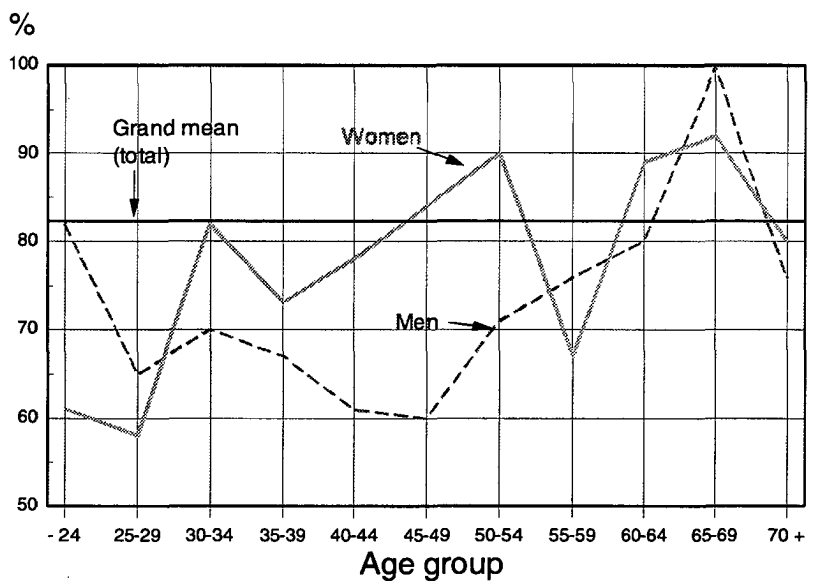


Figure 3 shows the percentage of traced telephone numbers in major cities in Finland. Helsinki, by far the biggest city in Finland, has one of the lowest rates of found numbers in the country. Only 45 of the more than 450 municipalities in Finland had lower rates than Helsinki, including only four very small towns. Two other major towns, Pori and Kuopio, (see figure 3) also had a very high proportion of non-telephone households. It is noteworthy that in these two towns there were no secret numbers.

Figure 4 shows the gender-age variation on telephone numbers in Helsinki. Young adults, especially women, appeared to have a very low rate of accessible telephone numbers. Another significant feature is the very low percentage of telephones of middle-aged men.

4. Discussion

What might be the reason for the difference between the telephone coverage estimated from The Household Survey (94 %) and the fact that a telephone number can be traced for only approximately 85 % of a random sample? Unfortunately this study cannot give a reliable answer to the question because of the limited number of variables analysed. However, some guidelines can be set.

The first cause for suspect, of course, is the method. Is the applied search method reliable enough? That is, are there many persons who actually can be reached by telephone, but their number could not be traced in telephone catalogues. This was checked and only a few additional numbers could be found by other methods. That is, the method explains only a small portion of the gap. On the other hand, the non-response rate is considerable in the Household Survey and in addition the wording does not clearly point to people's accessibility by telephone. Evidently there is reason to believe that the 'real' telephone coverage is somewhere between 85 % and 94 %.

This small-scale study showed that accessibility by telephone is connected to, e.g. a person's age, gender, and place of residence. However, there are probably still other characteristics connected to the ownership of a telephone that could not be treated here.

Smith (1990) has shown that in the United States low family income is the most important predictor for not having a telephone, along with other related socio-economic factors such as low education. According to many studies (e.g. Groves 1989, Smith 1990, Keeter 1995) there are many cultural and lifestyle variables that are significant as well. For instance, one-person households have been disproportionately represented among those with no phones, as were very large households.

Smith (1990) pointed out two different subpopulations who have fewer telephones than the rest of the population: outsiders with a weak attachment to society and its processes and institutions; and people in transition, e.g. because of recent separation, or recent or forthcoming moving.

Is this phenomenon typical only of Finland? If we disregard the magnitude of the non-telephone population, I would say not. Naturally, there are differences in the general levels of telephone coverage. However, the basic principles regulating the subscription to telephone services are

probably similar in most industrialised countries. For instance, Groves (1989) and Keeter (1995) have reported similar results in the United States.

If most of the people had a telephone and many of them could not be traced in telephone catalogues, then Random Digit Dialling (RDD) might be the solution. A prerequisite for this would be that people really have more telephones than can be found here. However, that does not seem to be the case. And according to Brick et al. (1995) the populations covered by RDD and not covered by RDD in the US show some differences, as well.

Another point is that the application of RDD techniques puts an end to the use of the advanced sampling theory because quota sampling would be the probable alternative for obtaining samples. Probability sampling enables the application of the whole arsenal of the sampling theory. The decision to give up the practice of drawing probability samples should be based on much sounder knowledge than is available at the moment.

The results lead still to yet an important question, that is, whether the telephone coverage obtained in the general purpose surveys is sufficient in measuring accessibility by telephone. The answer depends partly on specific local circumstances, but it seems that generally the actual telephone coverage, or accessability by telephone, should be defined in a specific survey.

At the moment, a specific survey on telephone coverage is even more important than a couple of years ago because telephone ownership and the entire concept of telephone coverage is changing due to the cellular phones. Their effect on telephone surveys is twofold: accessability will get better but, on the other hand, only short interviews are possible in cellular phones. In the future, the cellular phones will be a significant element in telephone surveys in many ways. For instance in Finland at the moment nearly 18 % of young people have only a cellular phone and the tendency is that more people and older people are giving up their ordinary phones.

References

Brick, J. M., Waksberg, J., Kulp, D. and Starer, A. (1995). Bias in List-Assisted Telephone Samples. *Public Opinion Quarterly*, 59, 218–235.

Groves, M. R. (1989). *Survey Errors and Survey Costs*. New York: John Wiley & Sons.

Keeter, S. (1995). Estimating Telephone Noncoverage Bias with a Telephone Survey. *Public Opinion Quarterly*, 59, 196–217.

Smith, T. W. (1990). Phone Home? An Analysis of Household Telephone Ownership. *International Journal of Public Opinion Research* 2, 369–390.

INTERVIEWER ATTITUDES AND UNIT NONRESPONSE IN TWO DIFFERENT INTERVIEWING SCHEMES

Risto Lehtonen
Statistics Finland, Finland

In the 1995 Finnish Health Security Survey, the data in the main survey was collected by professional interviewers of Statistics Finland and in a smaller-scale comparative survey, by public health nurses of local health centres. The response rate was 88 % in the main survey and 74 % in the comparative survey. There appeared to be significant differences in attitudes of interviewers towards the role of the interviewer between these two different interviewing schemes. The effect of the segregation of attitudes to nonresponse was, however, not substantial. The results are important for survey design because reluctancy to participate in surveys is increasing.

Key words: Finnish Health Security Survey, interviewer effect, nonresponse modelling, unit nonresponse.

1. Nonresponse in the Finnish Health Security Survey from 1964 to 1987

Five nationwide health security surveys have been conducted since 1964 by the Social Insurance Institution of Finland (Kalimo et al. 1992). In the first four cross-sectional surveys, all family members of samples of 6000–7000 households were interviewed in respondents' homes with PAPI method (paper and pencil interviewing) by public health nurses of local health centres. To collect the survey data in about three weeks, a large number of interviewers were needed. A total of 1800 nurses were attached to the interviewing task in each survey. All the interviewers were females. A field organisation, based on local health centres, was also established for each survey for the training of interviewers and the supervision of field work.

Overall response rates in the first four surveys are summarised in Table 1.

Response rate in the 1987 survey was quite high in other parts of the country except in greater Helsinki where the rate was 74 %. In other parts of Finland, the response rate was noticeably higher in rural regions than in urban regions (Lehtonen 1992).

Table 1. Response rate (%) in the Finnish Health Security Survey from 1964 to 1987.

Year	Interviewed households	Response rate %
1964	6400	91
1968	6900	97
1976	6700	91
1987	5800	84

2. Execution of the 1995 Finnish Health Security Survey

The 1995 Finnish Health Security Survey was executed in cooperation with the Social Insurance Institution of Finland and the National Research and Development Centre for Welfare and Health. For practical and economical reasons and to obtain high data quality, data collection for this survey was purchased from a professional interviewing organisation. The interviewing organisation of Statistics Finland was chosen for this purpose. In this interviewing organisation a total of 131 interviewers were available, of whom 127 were females. By using the interviewing organisation of Statistics Finland it was possible to collect the survey data using CAPI (computer-assisted personal interviewing), whereas in the earlier surveys PAPI was used. The data collection in the Finnish Health Security Survey was accomplished in two phases: 1800 households were interviewed in the spring of 1995, and 4200 additional households in the spring of 1996, producing a total of 6000 sample households to be interviewed. The survey data from these two data collection phases will be combined for analytical and reporting purposes.

Overall response rate in the first-phase survey in the spring of 1995 was 88 %. The rate was high, 85 %, also in greater Helsinki. In some rural regions the rate was over 92 %.

Several changes were made in the execution of the 1995 survey compared to the earlier surveys. The most important change was the use of professional interviewers who were not professionals in health care, whereas in the earlier surveys, public health nurses were used who were not professionals in survey interviewing. Therefore, a comparative survey was conducted simultaneously with the main survey to examine the possible effects of the changes in the data collection on the data comparability with the earlier surveys.

3. Experimental Design for the Comparative Survey

For the comparative survey, a total of 100 public health nurses were allocated for interviewing of a separate sample of 553 households with PAPI. Households were sampled from separately chosen 21 municipalities in Southern and Central Finland. The selected regions were survey regions also in the main survey and they were chosen so that desired variation with respect to certain demographic and socio-economic factors was achieved between the respective populations. A field organisation was also established, similar to the earlier surveys, for the training of interviewers and the supervision of the field work. In the comparative survey and the main survey, we attempted to make the necessary field operations reasonably well standardised, including contact processes, technical training, training in subject area, and supervision of field work. Similar material incentives were applied in both surveys.

Overall response rate in the comparative survey was 74 %. Regional response rates were also much lower than those of the corresponding regions of the main survey (Table 2).

Table 2. Regional response rates (%) in the main survey and comparative survey.

Region	Main survey	Comparative survey
Southern Finland	85	71
Central Finland	91	80
Overall	88	74

4. Attitudes Towards the Role of the Interviewer

There were relatively large overall and regional differences in nonresponse rates between the main survey and the comparative survey. One source of this difference was expected to be found in differences in professional background, and thus in interviewing skills and attitudes towards interviewing, between the two groups of interviewers. To examine this more closely, a questionnaire survey was conducted among the interviewers of the main survey and the comparative survey after having first completed the field work of both surveys. A total of 217 interviewers responded, 122 (93 %) in the main survey and 95 (95 %) in the comparative survey.

In this case study, differences in attitudes towards the role of the interviewer between the two groups of interviewers were examined. The following set of attitude questions, measured on a 5-point Likert scale, were included in the interviewer questionnaire (Question 23):

To what extent do the following arguments correspond to your own understanding about the role of interviewer?

Scale: 1 Strongly agree, 2 Agree, 3 Undecided, 4 Disagree, 5 Strongly disagree

	Agree ... Disagree
A. Reluctant respondent should always be persuaded to participate	1 2 3 4 5
B. With enough efforts, even the most reluctant respondent can be persuaded to participate	1 2 3 4 5
C. An interviewer should respect the privacy of respondent	1 2 3 4 5
D. If respondent is reluctant, refusal should be accepted	1 2 3 4 5
E. Voluntariness of participation should always be emphasised	1 2 3 4 5

For professional interviewers of the main survey and public health nurses of the comparative survey, the following results on the attitude questions were obtained:

Table 3. Proportion (%) of interviewers agreeing or strongly agreeing with the arguments A–E.

	Professional interviewers	Public health nurses
C. An interviewer should respect the privacy of respondent	96	99
E. Voluntariness of participation should always be emphasised	35	87
D. If respondent is reluctant, refusal should be accepted	27	82
A. Reluctant respondent should always be persuaded to participate	60	25
B. With enough efforts, even the most reluctant respondent can be persuaded to participate	29	15

The fundamental right to privacy (argument C) is widely accepted among both groups of interviewers as expected. The argument E, a general statement of emphasising voluntariness of participation, deals also with research ethics and is much more often agreed among public health nurses than among professional interviewers. The other three arguments measure attitudes towards persuading reluctant respondents. The argument D, stressing acceptance of refusal if the respondent is reluctant, and the reverse argument A, can be taken as special cases of argument E. As for argument E, clear differences can be seen in attitudes of the two groups of interviewers towards the arguments D and A. Among public health nurses, an attitude to agree with the argument of approving refusal of a reluctant respondent is much more often accepted than among professional interviewers. Finally, the difference between professional interviewers and public health nurses is not great in attitudes towards the argument B.

To more closely examine the differences in attitudinal orientation between professional interviewers and public health nurses, principal

components were constructed from the set of these five attitude items (scales of arguments A and B were first reversed). The first principal component explained 42 % of the total variation of the attitude measurements. For further analysis, this component was taken as a general index measuring attitudinal orientation towards the role of the interviewer. In this attitude index, low score indicates agreement with the attitude arguments (with scales of A and B reversed) and high score indicates disagreement.

Correlations of the original measurements with the general attitude index are shown in Table 4.

Table 4. Correlations of the original measurements with the general attitude index.

	General attitude index
D. If respondent is reluctant, refusal should be accepted	0.58
E. Voluntariness of participation should always be emphasised	0.52
A. Reluctant respondent should always be persuaded to participate(*)	0.46
C. An interviewer should respect the privacy of respondent	0.36
B. With enough efforts, even the most reluctant respondent can be...(*)	0.23

(*) Original scales were reversed for principal component analysis.

Largest correlations can be found in items D, E and A measuring the strength of attitudes towards reluctancy, voluntariness and persuasion. On the general attitude index, interviewers with low score tend to favour high voluntariness and low persuasion. Interviewers with high score, by contrast, tend to favour low voluntariness and high persuasion.

The 217 interviewers were further divided into three nearly equal-sized classes with respect to their score on the general attitude index. The following table indicates the relative position of the two groups of interviewers on the "High voluntariness, low persuasion – Low voluntariness, high persuasion" scale.

Table 5. Distribution (%) of interviewers on the general attitude index.

Voluntariness: / Persuasion:	High / Low	Medium / Medium	Low / High	Total
Professional interviewers (Main survey)	7	39	54	100
Public health nurses (Comparative survey)	68	27	5	100
All	34	34	32	100
Number of interviewers	73	73	71	217

By Table 5, more than a half of the professional interviewers of the main survey pertain to the "Low voluntariness, high persuasion" class, whereas among the public health nurses of the comparative survey the proportion is only 5 %. As a mirror picture, two-thirds of the public health nurses pertain to the "High voluntariness, low persuasion" class, while in the professional interviewers the proportion is less than 10 %.

Differences between the two groups of interviewers on the general attitude index were studied in more detail by logistic regression. Because the interviewers of the comparative survey were younger than those in the main survey (mean ages were 41 and 49 years, respectively), age was incorporated in the model to adjust for the possible age effect. Two binary response variables were constructed to study the variation in the probability of attitudinal orientation of the two extreme classes in Table 5. A response variable was constructed with a value 1 for interviewers in the "Low voluntariness, high persuasion" class and 0 otherwise, and another response variable for the other extreme in a similar way. Using these response variables each in turn, logistic regression was executed with SURVEY (0 for the main survey, 1 for the comparative survey) and the age of interviewer (as a five-score continuous variable) as the predictors.

Age adjusted odds ratios were estimated using the fitted logistic regression models. Age adjusted odds for "Low voluntariness, high persuasion" were 25 times higher for professional interviewers than for public health nurses, and age adjusted odds for "High voluntariness, low persuasion" were 29 times higher for public health nurses than for professional interviewers.

Estimation results are further illustrated in Figures 1 and 2. In Figure 1, the estimated probability of "Low voluntariness, high persuasion" is plotted against the age of interviewer, for both groups of interviewers. In Figure 2, this is done for the predicted probability of "High voluntariness, low persuasion". Figure 1 indicates that the estimated probability of "Low voluntariness, high persuasion" is much higher for professional interviewers than for public health nurses, in all age groups. For professional interviewers, however, the probability clearly decreases with increasing age. According to the mirror picture in Figure 2, the estimated probability of "High voluntariness, low persuasion" is much higher in public health nurses than in professional interviewers. The probability increases as the age increases in both groups of interviewers, but more apparently for public health nurses.

5. Unit Nonresponse and Attitude Segregation

A further question arises: does the segregation in attitudes towards the role of the interviewer among the two groups of interviewers affect the nonresponse figures on household level? This was examined by constructing a nonresponse model for the household-level sample of a total of 2353 households. Of these, 1800 households were from the main survey and 553 from the comparative survey. Logistic regression was used with a variable NONRESP (1 if a household is nonrespondent, 0 otherwise) as the binary

Fig. 1. Estimated probability of "Low voluntariness, high persuasion" by the age of the interviewer in the main survey (professional interviewers) and the comparative survey (public health nurses).

LOW VOLUNTARINESS, HIGH PERSUASION

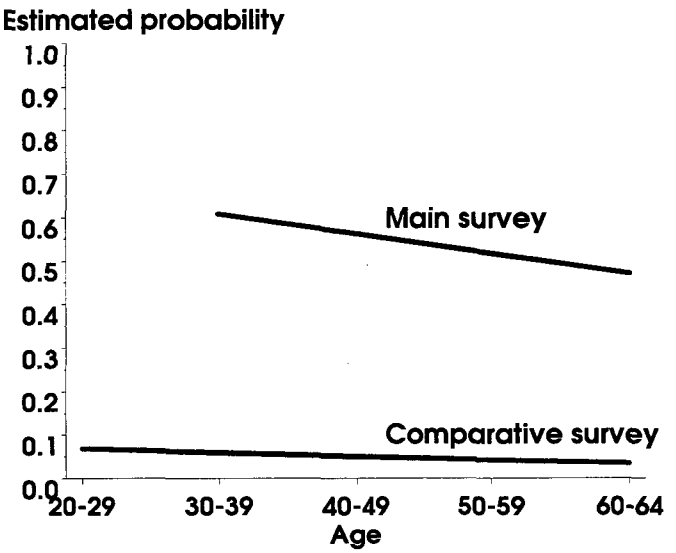
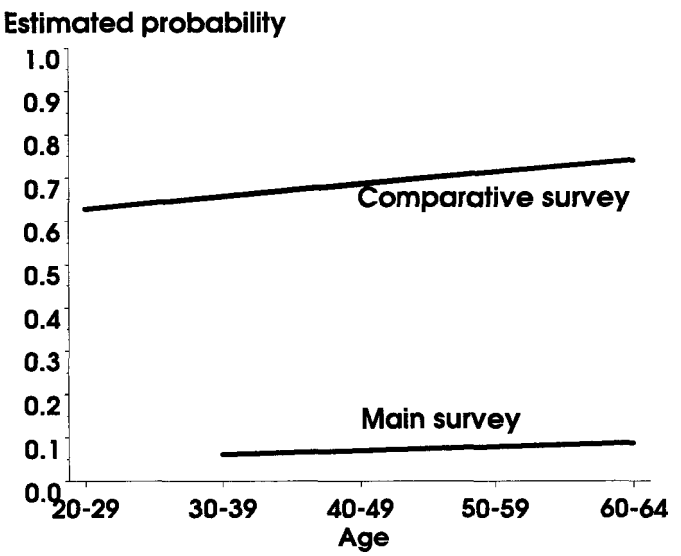


Fig. 2. Estimated probability of "High voluntariness, low persuasion" by the age of the interviewer in the main survey (professional interviewers) and the comparative survey (public health nurses).

HIGH VOLUNTARINESS, LOW PERSUASION



response variable and the age of interviewer (a five-score continuous variable) and the attitudinal orientation of interviewer as the predictors. For the first model, the interviewer attitude of the type "Low voluntariness, high persuasion" was taken as the attitudinal predictor (the variable HARDNESS: 1 if the attitude is present, 0 otherwise), and for the second model, the attitude of the type "High voluntariness, low persuasion" was used (the variable SOFTNESS: 1 if the attitude is present, 0 otherwise).

In model fitting, the fact that the households were clustered by interviewer (there were a total of 217 such clusters) was taken into account by using design-based methods to obtain valid estimation and testing results (Lehtonen and Pahkinen 1995). More specifically, the first-order version of the generalised estimating equations technique (GEE; Diggle et al. 1994) with an exchangeable correlation structure was used with a home-made SAS macro. Alternatively, for example the SUDAAN procedure LOGISTIC (Shah et al. 1995), aimed at design-based logistic regression, could be used for this purpose as well.

The results from logistic regression are summarised in Table 6. There, results from model-based estimation that ignores the clustering of households, are given for comparison with the more reliable results from design-based estimation where the clustering is appropriately accounted for.

Table 6. Design-based and model-based estimates of beta coefficients of the fitted logistic regression models with the accompanying standard errors and design effects, and observed values and the corresponding p-values of a Wald test statistic.

	Beta coefficient	Standard error	Design effect	Wald chi-square	p-value
Hardness					
Design-based					
intercept	–0.86	0.214	1.3	16.0	0.0001
age	–0.27	0.067	1.2	15.9	0.0001
hardness	–0.09	0.134	1.3	0.5	0.4785
Model-based					
intercept	–0.86	0.195	1.0	19.3	0.0001
age	–0.27	0.062	1.0	18.3	0.0001
hardness	0.11	0.123	1.0	0.7	0.3926
Softness					
Design-based					
intercept	–1.01	0.232	1.4	19.0	0.0001
age	–0.26	0.069	1.4	13.6	0.0002
softness	0.31	0.153	1.4	4.0	0.0452
Model-based					
intercept	–1.03	0.205	1.0	25.2	0.0001
age	–0.25	0.062	1.0	15.9	0.0001
softness	0.32	0.135	1.0	5.7	0.0175

The results indicate that when adjusting for the age of interviewer, the probability of being a nonrespondent household does not depend on the interviewer attitude of the type "Low voluntariness, high persuasion" i.e. HARDNESS. On the other hand, the probability does depend weakly on the interviewer attitude of the type "High voluntariness, low persuasion" i.e. SOFTNESS. The probability of nonresponse slightly increases as this attitudinal orientation increases. For both models, the age effect is statistically significant: the probability of nonresponse decreases as the age of interviewer increases.

It should be noted that for the actual clustered design, design-effect estimates are uniformly larger than one indicating that clusters of households (i.e. interviewer workloads) tend to be internally homogeneous with respect to nonresponse.

6. Discussion

The properties of survey interviewers play an important role as regards survey errors (Groves 1989; Brehm 1993). It was demonstrated in this case study that interviewer's attitudes towards the role of the survey interviewer can be an important source of nonresponse in a survey based on face-to-face interviewing. In the main part of the Finnish Health Security Survey, where professional interviewers were used, an overall response rate of 88 % was attained. The rate was 74 % in the additional comparative survey, where health care professionals (public health nurses) were used as interviewers. Attitudes towards the role of the interviewer differed substantially between the two groups of interviewers, as appeared in an interviewer questionnaire conducted after completing the field work of both surveys. Attitudes of interviewers were measured by two Likert scale questions concerning privacy and voluntariness as general concepts of research ethics, and by three more specific Likert scale questions concerning the behaviour of the interviewer when working with reluctant respondents.

Experienced professional interviewers such as those in the main survey often are equipped with high skills for motivation and persuasion of respondents which skills are needed especially when working with reluctant respondents. In survey interviewing practice, these interviewer properties appear to purposefully aim at interviewing all kinds of respondents including the most reluctant ones. On the contrary, public health nurses who worked as interviewers in the comparative survey were not professionals in survey interviewing but were professionals in health care. Public health nurses had also good skills for face-to-face communication with people but their attitudinal orientation towards the role of the survey interviewer appeared to differ substantially from that of professional interviewers. When

compared with professional interviewers, public health nurses more often tended to emphasize voluntariness of participation, leading to a tendency of accepting refusal without further persuasion when respondent appeared reluctant.

A general attitude index was constructed using the set of the five attitude measurements by taking the first principal component of the corresponding variables. Scores on the resulting attitude index tended to be higher among professional interviewers than among public health nurses indicating in professional interviewers an attitudinal orientation towards an interviewer role that favours low voluntariness of participation and high persuasion of reluctant respondents. In public health nurses, an attitudinal orientation towards the role of the interviewer that favours high voluntariness of participation and low persuasion of reluctant respondents was noticed.

Professional interviewers and public health nurses thus differed substantially with respect to their attitudinal orientation towards the role of the interviewer measured by the general attitude index. More than a half of professional interviewers possessed a "hard" attitudinal orientation favouring low voluntariness and high persuasion. This attitudinal orientation, however, varied among professional interviewers so that it was more common among young interviewers than among old interviewers. This may indicate that with increasing interviewing experience, more sophisticated and individual-oriented methods are applied when working with reluctant respondents. Also a slight tendency towards "soft" attitudes favouring high voluntariness and low persuasion as age increased was noticed among this group of interviewers.

Two-thirds of public health nurses possessed an attitudinal orientation favouring high voluntariness and low persuasion, whereas only 5 % favoured low voluntariness and high persuasion. One possible explanation to this segregation might be in the ethical norms guiding everyday practice of public health nurses when working at local health centres. Rules in their interaction and communication with people lay upon medical ethics that can be more obligatory e.g. with respect to privacy protection and confidentiality than are ethical rules prevailing in typical survey interviewing. Norms of medical ethics perhaps guided the behaviour of public health nurses also when they worked as survey interviewers. With increasing professional experience in health care, adopting "soft" attitudes towards the role of the interviewer tended to increase among this group of interviewers.

When examining the nonresponse at household level, it appeared that the probability of nonresponse did not vary statistically significantly between "hard" type and "non-hard" type interviewers. On the other hand, the probability did vary somewhat between "soft" type and "non-soft" type interviewers so that the probability slightly increased as "soft" attitudinal orientation increased. The age of the interviewer was statistically significant:

the probability of nonresponse decreased with increasing age. Clusters of households (i.e. interviewer workloads) tended to be to some degree internally homogeneous with respect to nonresponse.

Reluctancy to participate in surveys is increasing also in Finland, although relatively high response rates can still be attained as was the case in the main part and the comparative part of the Finnish Health Security Survey. When attempting to keep nonresponse at an "acceptable" level, the important problem of appropriate behaviour of interviewers should, however, be considered in connection with the quality of collected survey data. Research on this problem is presently carried out in the research group of the Finnish Health Security Survey.

Acknowledgements. I am grateful to Sirkka-Sisko Arinen, Timo Klaukka and Tarja Nieminen, all of the Social Insurance Institution of Finland, for constructive comments.

References

Brehm, J. (1993). *The Phantom Respondents. Opinion Surveys and Political Representation.* Ann Arbor: The University of Michigan Press.

Diggle, P. J., Liang, K-Y. and Zeger, S. L. (1994). *Analysis of Longitudinal Data.* Oxford: Oxford University Press.

Groves, R. M. (1989). *Survey Errors and Survey Costs.* New York: John Wiley & Sons.

Kalimo, E., Klaukka, T., Lehtonen, R., Nyman, K. and Raitasalo, R. (1992). Health care development in Finland, 1960 to 2010. *World Health Forum* 13, 336–342.

Lehtonen, R. (1992). Terveysturvan väestötutkimuksen aineistot ja tutkimusmenetelmät. Part IV in Kalimo, E., Klaukka, T., Lehtonen, R., and Nyman, K. *Suomalaisten terveysturva ja sen kehitystarpeet.* (Summary: *Health security in Finland and needs for development*). Helsinki: Publications of the Social Insurance Institution, Finland, M:81.

Lehtonen, R. and Pahkinen, E. (1995). *Practical Methods for Design and Analysis of Complex Surveys.* Chichester: John Wiley & Sons.

Shah, B. V., Barnwell, B. G. and Bieler, G. S. (1995). *SUDAAN User's Manual: Software for Analysis of Correlated Data, Release 6.40.* Research Triangle Park, NC: Research Triangle Institute.

NONRESPONSE RESEARCH PLANS FOR THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

Stephen P. Mack and Preston J. Waite
U.S. Bureau of the Census, U.S.A.

Nonresponse bias is an important concern for longitudinal surveys such as the Survey of Income and Program Participation (SIPP). Previous panels started with nonresponse rates around 7.5 % in the first wave and climbed to around 20 % by the eighth wave. We are making changes to the 1996 panel that are expected to increase nonresponse rates: the panel length will increase to four years, we will introduce computer assisted personal interviewing (CAPI), and the panel size will be larger than previous panels (so many interviews will not have previous experience with the SIPP). The U.S. Census Bureau has initiated several projects with the goal of reducing nonresponse bias. This paper briefly describes these projects.

Key words: Imputation, weighting, incentives.

1. Introduction

The SIPP interviews sample households every four months over a period of years. The level of household nonresponse increases cumulatively with each interview. We are making changes to the 1996 panel that are expected to result in increased nonresponse rates.

- The 1996 panel will have 12 waves rather than 8. This will increase the length of the panel to 4 years.
- We will begin computer assisted personal interviewing (CAPI). Other Census Bureau surveys have experienced increased nonresponse rates when CAPI was started. The SIPP used CAPI in a 1995 dress rehearsal which had wave 1 nonresponse rates of around 12 %.
- The size of 1996 panel will increase to approximately 37,000 interviewed households. As a result, many interviewers will not have had previous experience with the SIPP.

The longer panel will improve the utility of SIPP data for longitudinal analysis, but nonresponse bias will be a concern. Sampling variances of estimates will be smaller due to the larger panel size. The accuracy of some statistics, for example the poverty rate, might not improve if nonresponse bias increases too much.

The U.S. Census Bureau has initiated several new research projects with the goal of measuring and reducing nonresponse bias. The goals of the projects fall into the three general categories below.

- Measurement of nonresponse bias
- Reduction of nonresponse
- Improvements to nonresponse adjustment methodology

The specific projects are discussed in the next section.

2. Nonresponse Projects

Examining Nonresponse/Attrition by Field Representatives

This project will gather information from field representatives (FR's) and supervisory field representatives (SFR's) about effective methods of obtaining interviews, finding movers, and converting refusals to interviews. FR's and SFR's who historically maintain high response rates will be asked to have methods they use to find movers and convert refusals. We hope that analysis of this information will provide insights on how training materials and data collection procedures can be improved.

We plan to use information from this project in interviewer training for the 1996 panel.

Re-assessing how we Inform Respondents and Interviewers about the Importance of the Survey

The purpose of this project is to better motivate respondents and interviewers about the importance of the SIPP.

We've been told by our regional offices that we need to do a better job of publicizing the value of the SIPP to the nation and local communities. This project will examine our current practices of publicizing the SIPP and that of other surveys. The expected results are:

- Recommendations about how to improve the way we use our current materials
- Ideas for new ways to publicize the SIPP and inform sample households about why it is important that they respond.

Nonresponse Follow-up Study

Household nonresponse in wave 1 (the first interview) of the SIPP averaged about 7 % for the 1984 thru 1991 panels. The two most recent panels (1992 and 1993) had average nonresponse rates over 9 % in wave 1. Since the SIPP is a longitudinal survey, nonresponse rates are cumulative. Households that do not respond in wave 1 or in two consecutive later waves are not interviewed for the remainder of the panel.

Field representatives try to obtain some information about wave 1 noninterviews, such as race and tenure, for use in nonresponse adjustment. Additional information would allow us to better understand the characteristics of wave 1 nonrespondents and seek ways to reduce nonresponse bias. We would like to evaluate:

- Differences of distributions of wave 1 respondents versus nonrespondents for variables such as income, poverty, and program participation.
- The effectiveness of the current nonresponse adjustment in reducing bias
- Alternative nonresponse adjustments

Plans for this study call for a short questionnaire to be mailed to all wave 1 nonrespondents. Nonrespondents will be asked to fill in basic information on income, program participation, and household membership. Information from returned questionnaires will be used to prepare a new noninterview adjustment and weights. We will use the alternative and original weights to form two set of key estimates. The ability of the nonresponse follow-up to reduce bias will be evaluated by comparing the two sets of estimates. In addition, we will compare the differences between respondent and nonrespondent populations.

Accuracy of "Type Z" Imputations

We call noninterviews of persons within interviewed households a type Z noninterview. Households in which at least one person was interviewed are interviewed households. Type Z nonresponse is the second most prevalent type of nonresponse at the person level in waves 2 and later. About half of wave 1 type Z cases are never interviewed during the panel.

We use hot deck imputation to fill in data for Type Z noninterviews. Type Z cases are matched to donors and data from the donor case is substituted for the missing interview. Other types of nonresponse are generally handled by weighting adjustments.

In this project, we plan to study:

- Alternative field procedures to reduce the amount of type Z nonresponse
- The effect of type Z nonresponse on key estimates

The first part will consist of reviewing of field procedures. Changes to proxy rules and other procedures will improve the chances of collecting at least some information for all household members.

The second part will consist of an evaluation of the hot deck procedure. Additional type Z's will be created from fully reported cases and imputed for missing data. Key statistics, such as poverty rates, will be retabulated and compared with the original estimates. If we find significant differences on the key statistics, we will investigate and evaluate alternative imputation procedures.

Table 1. Percent of nonresponse of persons in 1991 panel by type of nonresponse.

Wave	Type Z	Whole Household Refusal	Mover	Other
2	34.1%	39.1%	14.8%	12.0%
3	30.3%	42.0%	19.6%	8.2%
4	29.2%	41.8%	23.0%	6.0%
5	28.6%	39.9%	24.0%	7.5%
6	27.6%	40.0%	26.6%	5.8%
7	30.1%	38.2%	26.3%	5.4%
8	30.7%	36.7%	27.7%	4.8%

Carry-over Imputation

The SIPP provides longitudinal weights for the analysis of persons over a particular calendar year or over the panel. Persons who are classified as interviewed (self, proxy, or imputed) for the appropriate period receive a positive weight. Starting with the 1991 panel, we began using data from surrounding interviews to fill in data for a missing interview. A random procedure is used to determine the number of months (0 to 4) that will be imputed with data from the last month before the missing wave, and the number of months imputed the data from the first month after the missing wave. We restrict this procedure to filling in missing interviews that are bounded on two sides by a self or proxy interview. The 1990 panel was used to test carry over imputation. The impact of carry over imputation on the number of positively weighted cases is shown below.

Table 2. Increase in positively weighted persons in 1990 panel due to carry-over imputation.

	Increase in positively weighted cases	
Weight	Number	Percent of Designated Sample
Panel	3521	5.7%
1990 Calendar Year	2450	4.0%
1991 Calendar Year	2095	3.1%

The number of positively weighted cases can be increased by imputing interviews for other nonresponse patterns. For example, we can increase the number of positively weighted cases about 1 % by imputing for 2 consecutive noninterviews bounded by interviews.

This study will explore the merits of extending carry-over imputation to bounded and nonbounded noninterview patterns of up to two waves. Extending carry-over imputation may increase bias in some estimates, such as number of transitions into and out of poverty. We will weigh the benefits of additional positively weighted cases against possible increases in bias.

Incentives Test

Past research has shown that incentives are effective in increasing response rates. A calculator incentive was tested in the SIPP 1987 panel (Butler 1991). The calculator incentive resulted in a 2 % increase in final response rates compared to the control (non-incentive) group. Another study (Willimack 1995) offered a nice ball point pen, which increased response rates from 76 % to 81 %. Other studies have shown that the number of callbacks needed is reduced when incentives are given, which decreases interviewing costs.

The incentives test will be conducted during wave 1 of the SIPP 1996 panel. Three treatments are proposed:

Control	No incentive
Treatment 1	$10
Treatment 2	$20

Field representatives will distribute the incentive to sample households prior to the first interview. The experimental design calls for stratifying PSU's according to size into 3 strata. Within each strata, treatments will randomly be assigned to PSU's. The incentives test is expected to answer the following:

- Do incentives significantly increase wave 1 response rates?
- Do incentives in wave 1 increase response rates in later waves?
- Do incentives reduce the number of callbacks?
- Does the effect of the incentive depend on the value of the incentive?

Effect of Second Stage Weighting on Nonresponse Adjustment

SIPP cross-sectional (and longitudinal) weights are the result of a series of adjustments to designed to make the distribution of sample characteristics more closely resemble the general population. The components of cross-sectional weights are listed below.

- **Base Weight (BW)** – The inverted probability of selection of a person's household.
- **Duplication Control Factor (DCF)** – Adjusts for subsampling done in the field.
- **Wave 1 Noninterview Adjustment Factor (F_{N1})** – Adjusts for non-interviewed households in wave 1.
- **Movers Weight (MW)** – Adjusts for persons in the SIPP universe who move into sample households after wave 1.

- **Wave 2+ Noninterview Adjustment Factor (F_{N2})** – Adjusts for noninterviewed households in waves 2 and later that were interviewed in wave 1.
- **Second Stage Adjustment Factor (F_{2S})** – To adjust estimates to population controls.

The second stage adjustment is large for some populations. For example, the number of black households increased by 14.3 % in the March 1991 second stage adjustment. Black households in poverty increased even more, 22.3 %.

It is possible that effectiveness of nonresponse adjustment is reduced by the second stage weighting adjustment. A few reasons why this seems plausible are:

- We have investigated a number of alternative wave 2+ nonresponse adjustments. After each alternative, second stage adjustment factors were computed to obtain final weights. Estimates of key statistics using the original weights and alternative weights were all very similar to each other.
- In raking adjustments of two or more marginals, the sum of weights corresponding to a given marginal will not agree with the controls of that marginal after raking on other marginals. This situation is comparable to the situation in the last two stages of weighting.

We will investigate the effect of the second stage adjustment on the nonresponse adjustment by computing alternative weights. The alternative weights will be obtained by starting with wave 1 final weights and then doing the last three stages of weighting. Second stage factors are expected to be smaller (and effect the nonresponse adjustment less) under the alternative procedure than they are under the original procedure. Estimates of key statistics will be compared to evaluate the importance of any effect. If differences are found, we will investigate weighting alternatives with combined nonresponse and second stage adjustments.

3. Conclusion

The long length of the 1996 SIPP panel will make nonresponse an important concern throughout the panel. Wave 1 nonresponse is particularly important since wave 1 nonrespondents are nonrespondents for the entire panel. Three projects focus primarily on wave 1.

- Examining nonresponse/attrition by field representatives.
- Incentives Test.
- Nonresponse follow-up study.

The first project looks for ways to improve interview training; the second project will study the effect of incentives distributed in wave 1 on response

rates; and the third project will collect information on wave 1 nonrespondents.

We have initiated new research projects that seek to reduce nonresponse bias by:

- characterizing nonrespondents and measuring nonresponse bias
- reducing nonresponse rates
- and improving nonresponse adjustment methodology.

We are also continuing to work on other investigations:

- Regression weighting methods (An, Breidt, and Fuller 1994).
- Using Internal Revenue Service (IRS) income data to improve weighting (Dorinski and Huang 1994).

References

An, A., Breidt, F. and Fuller, W. (1994). Regression Weighting Methods for SIPP Data. *American Statistical Association 1994 Proceedings of the Section on Survey Research Methods.*

Butler, D. (1991). SIPP 1987: Gift Experiment Results. Internal Census Bureau Memorandum from Butler to Singh, April 2, 1991

Dorinski, S. and Huang, H. (1994). Testing a New Attrition Nonresponse Adjustment Method for the SIPP. *American Statistical Association 1994 Proceedings of the Section on Survey Research Methods.*

Willimack, D., Schuman, H., Pennell, B. and Lepkowski, J. (1995). Nonmonetary Incentives in Face-to-Face Surveys. *Public Opinion Quarterly*, 59, 78–92.

NONRESPONSE IN THE HUNGARIAN HOUSEHOLD SURVEYS

Adam Marton
Central Statistical Office
Hungary

In household surveys, which are face-to-face interviews, the response rate is fairly high. This is due partly to the good sampling frame and field work and also to the long expertise. But the response rate differs depending on the topic of the survey and strata of the population.

The treatment of nonresponse is different in the case of various surveys. The paper describes the basic features of the sampling design, and how the required number of answers can be obtained. The aim is to reduce the sampling error and also the bias originating from different sources. Substitutes are often used, which practice can be criticized.

The field work organization has changed substantially from the beginning of this year. Up to now, it is obvious that this change has some positive and negative impact as well on the quality of the results.

Key words: Household surveys, nonresponse, substitution.

The Unified System of Household Surveys (USHS) has been operating since 1976. It is part of the Hungarian Central Statistical Office (CSO), and covers the noninstitutional population

There are two large ongoing household rotating sample survey based on face to face interviews:

- The Family Expenditure Survey (FES)
- The Labor Force Survey (LFS).

Supplementary, short questionnaires are sometimes added to the FES and/or LFS.

There are several other topics to be investigated, some of which are regularly repeated in 2, 5, or 10 year intervals. These are: survey of household income, social mobility, grading of occupations, time budget, living conditions of young and older generations, health behavioral risk factors, travelling habits, etc.

The sample size depends on the resources available and the required precision. The questionnaires are fairly complicated and contain many

questions. The total survey error of the variables observed in the different domains and regions is different.

The subject-matter statistitian's main and most important target is to get the required number of completed questionnaires, which can be criticized, from different aspects.

The whole sampling design, field work is coordinated by the subject matter statisticians and survey methodologists of the CSO.

The most common reasons of nonresponse:

- refusal
- partial refusal
- empty dwellings
- not at home
- not able to answer
- wrong address
- other

The interviewers' activity is organized and supervised by the County Offices of the CSO. The interviewers in the smaller towns and villages establish, in many cases, some kind of personal contact with the members of the households in the sample. This helps to reduce refusals, but as far as the response variance is concerned, it has a certain disadvantage too. The given arrangement of the field work is crucial: how to handle nonresponse, nonsampling error, etc.

Careful (follow-up) studies, call-backs help-to reduce the number of nonresponse and also try to get some basic information on the household, since the bias can be also reduced through poststratification.

It is obvious that the estimates for the "extremes" of the target population (very poor, very rich) are unreliable due to high nonresponse rate and also certain nonsampling errors.

There is a widely used method in Hungary and also in some other countries. In addition to the original probability sample another, additional probability sample is selected. Then in case of nonresponse in the first sample, through some prescribed way the missing sampling unit is to be substituted. In this way the desired sample size can be easily obtained.[1] This procedure reduces somewhat the sampling error, but not the bias, because the distribution of refusals is very likely to be the same as in the first sample.

Using substitutes is a special way of handling nonresponse. Popularity originates in the misbelief of some statisticians: the actual sample size is the crucial point as far as the quality of data is concerned. Nevertheless, this procedure may give some help.

---

[1] At the previous Workshop in Ottawa two papers (Kordos, Vehovar) discussed this problem.

It's advantages:

- the required sample structure can be stabilized,
- does not increase the bias if there is some information on the basic characteristics of the missing units.
- field work can be balanced.

Disadvantage also has to be considered:

- the interviewers do not make enough effort to complete the interview at the primary unit if they are faced with some difficulties, when it is allowed to replace them,
- it is not easy to control whether the interviewer strictly follows the prescribed rule.
- it increases the volume of workload, because more addresses are to be visited than the required sample size.

The LFS is a quarterly survey (the sample covers proportionally all months) and after 6 visits the household is replaced. The response rate – as will be shown later – is fairly high. (Since 1990 unemployment has became a serious problem in Hungary.) The questionnaire is a fairly simple one and at the consecutive visits only the changes have to be noted.

Table 1. Labor Force Survey.

			Number of			
	Total	**respon-dents**	**out of scope**	**empty dws**	**non existing**	**refu-sals**
Total	26683	22502	702	2269	370	840
percentage	(100)	(84)	(3)	(8)	(2)	(3)
Budapest	4038	2947	415	223	115	327
percentage	(100)	(73)	(10)	(6)	(3)	(8)

In this case it is supposed that the distribution of missing data is random, so the inflation rate is adjusted in each stratum. No substitutes are used.

The FES is burdensome for the households. They are asked to prepare a diary for one month containing all expenditures b by categories of commodities, services, etc. After the end of the year all households are visited again to get some information on the expenditures of "big" items (consumer durables, expensive clothing, reconstruction etc.) and also on the yearly incomes.

The survey procedure is as follows. A list of households is given to start with the interviews. If the address is wrong, empty, or the household refuses to participate, then from a supplementary list of addresses a new unit has to be selected till the required number of households is reached. If the nonresponse and the substitute units distributed randomly, there would be not be any problem related to the nonresponse. But this is not the case.

The refusal rate is very high, 30–40 %, due to the difficulties of preparing diaries. In this case, the bias coming from refusals is much more serious than that of the LFS.

Table 2. Family Expenditure Survey.

	Total	respon-dents	out of scope	Number of empty dws	non existing	refu-sals
	Diaries					
Total	13327	8111	747	452	516	3501
percentage	(100)	(61)	(6)	(3)	(4)	(26)
Budapest	2026	660	380	64	58	864
percentage	(100)	(33)	(19)	(3)	(3)	(43)
	End of the year interview (1994 March)					
Total	8106	7495	21	20	120	420
percentage	(100)	(92)	(0.2)	(0.2)	(2)	(5)
Budapest	660	541	7	1	12	99
percentage	(100)	(82)	-	~	(2)	(15)

1/ Five is lost 2/ Partial refusal (income) included.

In certain cases additional surveys are carried out using units of earlier LFS, FES or other surveys after they have been rotated out. But a given sample suffers some attrition after several visits. At the same time reassigning the already used element has the advantage of knowing the demographic characters of the given household and its members. Using further substitutes, the representativity of the sample can be maintained. This procedure needs further analysis. (There is an example when a given sample was used three times for three different surveys while the sampl was getting smaller and smaller.)

Since the beginning of this year the interviewers have been paid according to the actual workload. Earlier they were full or part time employees of the CSO. The experiences have to be scrutinized, but it seems to be clear, that the response rate has increased.

Finally it is very important to control field work using some kind of incentives, and the training of interviewers is also important.

Annex

Sample size and rate of nonresponse

Family Expenditure Survey (substitutes)

	Actual s.size	Rate of nrp
1983	7986	37.6
1985	11844	35.4
1987	9144	33.8
1989	8940	35.5
1991	8760	37.0
1993	8116	39.1
1994	7926 (13891)	43.0

Income Survey (substitutes)

1983	15780	9.4
1987	19820	17.3
1991	14474	26.0

Labor Force Survey (no substitutes)

1992	17.2
1993	17.B
1995. I-II.	18.4

RESPONDENTS VS NONRESPONDENTS: HOW DIFFERENT ARE THEY ?

Sylvie Michaud and Michel Latouche
Statistics Canada, Tunney's Pasture
Ottawa, Ontario K1A OT6
Canada

Longitudinal surveys are becoming more popular as a tool to help understand causal processes that may feed into policies. To ensure the validity of the results, great attention has to be paid to the response rate of the surveys, especially at the onset, since usually, only respondents to the first contact will be followed-up to become the longitudinal cohort. Statistics Canada implemented three longitudinal surveys in 1994. This paper looks at the response rates across the surveys, and does some comparison between respondents and non-respondents to try to get a better understanding at response mechanisms.

Key words: Differential non-response, administrative files, log-linear models.

1. Introduction

In 1994, Statistics Canada implemented three new longitudinal social surveys; the Survey of Labour and Income Dynamics (SLID), the National Population Health Survey (NPHS) and the National Longitudinal Survey of Children (NLSC). High response rates in a longitudinal survey are very important, especially at the onset, since the set of respondents from the first interview usually becomes the cohort of people that will be followed. In particular, when non-respondents are refusals, a balance has to be maintained between following rules, costs of following non-respondents, response burden and data quality. Section two will present the problem and the observed response rates. Section three will look at the field procedures and the conversion rates obtained in the different surveys, to increase response rate. However, even when people are converted from non-respondents to respondents, the issue remains are there different kinds of non-respondents (i.e. not converted vs converted) and is the conversion process compensating for a non-ignorable non-response mechanism. Some non-respondents were linked to an administrative file to get a partial answer

and the results are presented in section four. Finally, section five will present additional research that has been done on non-response and outline future research.

2. Response Rates

Impact of Non-Response on Estimates

As mentioned in Binder et al. (1994), let's consider a nonresponse adjustment which we refer to as *generalized reweighting methods*. In a sample s, a set of responding units s' is observed. Associated with each responding unit, k, we have an adjusted weight given by

$$w_k'(s',s) = g_k'(s',s)w_k(s)$$

where $g_k'(s',s)$ is a weight adjustment that makes use of auxiliary frame data, as well as other information that may be available for the nonresponding units. This allows the weight adjustment to depend on survey values that were observed on previous occasions from a longitudinal survey. We assume that the estimator of a total for a y-variable on the t-th occasion is given by

$$\hat{Y}_{it}^{(GR)} = \sum_{k \in s_t'} w_k'(s_t', s_t) y_{ikt} \qquad (2.1)$$

We let $\rho_k(s)$ be $\Pr(k \in s' | s)$. Assuming that the original estimator was consistent, (2.1) will be asymptotically consistent with respect to the original design and the response probabilities if:

1) the probability distribution of s' given s depends only on the auxiliary data and the survey data from previous occasions, but not directly on the y-values for the current occasion,
2) the limiting expectation of $g_k'(s',s)$ is $\{E[\rho_k(s)]\}^{-1}$,
3) the variance of $\hat{Y}^{(GR)}$ is asymptotically zero.

If condition 2) is violated, then the expectation of $\hat{Y}^{(GR)}$ is

$$\sum E[g_k'(s',s)]E[\rho_k(s)]y_k . \qquad (2.2)$$

The form of this bias is important, because if one were to impose model assumptions on the y-variables, it is possible that the model-bias becomes

small. However, for those who wish to make the fewest model assumptions, it is clear that one should restrict attention to adjustment methods which yield condition 2) as closely as possible. This implies that the weight adjustment should reflect the propensity to respond as nearly as possible. Of course, the probability mechanism generating these response probabilities are generally unknown, so the weight adjustment must necessarily be model-based.

Another important feature of 2) is that if there are some "hard-core" nonrespondents – that is, units where $\rho_k = 0$ – there would be no consistent estimates. More details can be found in Binder et al. (1994).

Response Rates for the Longitudinal Surveys

As mentioned earlier, in 1994 Statistics Canada implemented three longitudinal social surveys. The surveys share some characteristics. They are all sampled using a multi-stage sample design (using the frame of the Canadian Labour Force Survey (LFS)). They are also collected in a decentralised computer assisted interviewing mode. Finally, they share a common "pool" of interviewers. There are however differences between the surveys. SLID and NLSC samples were selected from people who previously participated in the LFS. People in the NPHS have never been contacted before. Most interviews in SLID are done by telephone. NPHS and NLSC collect their information with a personal visit. Figure 1 shows the response rates for the different surveys (before the follow-up operation). The LFS survey usually has a response rate around 95 %. The rates for SLID and NLSC have been adjusted to compensate both for the survey non-response as well as for LFS non-response. The rates are produced by regions in Canada. More details on the study of response rates for the longitudinal surveys can be found in Statistics Canada (1995).

NLSC has the lowest response rates. Table 1 shows it also has the longest interview. There are also regional differences for the surveys. It is not clear how much previous burden impacted on the response rate. Before the last follow-up from NPHS, SLID's response rates were in general a higher response rate than NPHS. However, SLID is shorter, and the contents are different. There are also confounding factors that make the comparisons difficult (like the number of surveys that are in the field at the same time). Other than noting that time (in a large sense) had an impact on the response rates, conclusions were limited. So we focused next on the extent of improvements to the response rates that are gained by a response follow-up.

Fig. 1. Cross-sectional response rates by region for the different surveys.

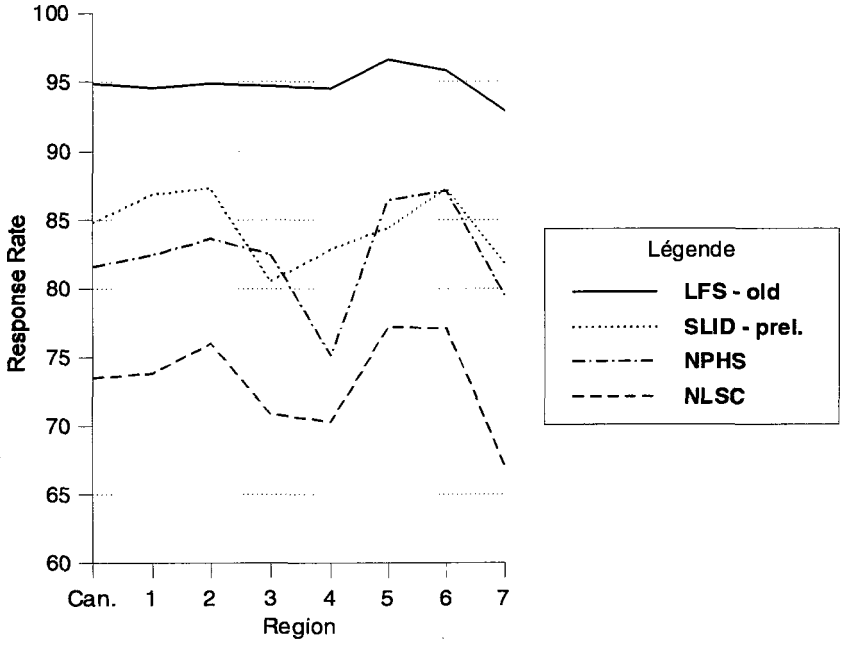


Table 1. Average length of interview and frequency of contacts.

	interview length (per household)	frequency of interview
LFS	7	once a month
SLID	20-30 minutes	twice a year
NPHS	75 minutes	every two year
NLSC	120 minutes	every two year

3. Impact of Follow-up on Response Rates

SLID's Follow-up

SLID's sample is a sub-sample of LFS respondents (LFS has a response rate of 95 % and one third of these respondents have been selected to be in SLID). The SLID's preliminary interview got a response rate of 88 % (the preliminary interview informs respondents they will be contacted for an extra six years, and asks for marital and work history). For costs reasons, only respondents were followed, since most non-respondents were refusals. However, for data quality evaluation purposes, two hundred refusal households were reinterviewed in SLID first year of interview. 63 % non-respondents participated in the first year of SLID, which suggest that some refusals could be converted to response.

NPHS Follow-up

NPHS collection was spread over a full year. Every three months, a quarter of the sample was interviewed. This design allowed the transfer non-respondents of a given quarter in the subsequent quarters for follow-up (a few people were excluded from this process when they expressed strong feelings about not being interviewed). A number of tables were produced to estimate the extent of conversion of non-respondents. As can be seen from table 2, sending back non-respondents in the first quarter gave more pay-off than in subsequent quarters. Similar results were found in quarter 2.

Table 2. Conversion rates for NPHS quarter 1.

	quarter 1	quarter 2	quarter 3	quarter 4
resp rates	80.6 %	86.3 %	87.3 %	88.2 %
increase		5.8 %	1.0 %	0.9 %

NLSC Follow-up

NLSC's sample collection is divided in two periods; one in December and one in February. The same transfer mechanism was applied to NLSC; non-respondents to December were transferred for the February collection. As can be seen from the second column of table 3, the increase in response rate was slightly lower than for NPHS. Response rates were felt to be still under the targets and a final separate follow-up was conducted in June. At that time, regional office managers were brought in, and more management tools were offered to them based on observed results. Based on NPHS findings, without special efforts, there was a minimum expected gain of 1 % of non-respondents by this second round of conversion. The second follow-up operation actually provided better results than expected with an increase in response rate of 6.0 %.

Table 3. Conversion rates for NLSC December collection.

	December collection	Feb. follow-up	June follow-up
response rates	73.4 %	77.0 %	83.0 %
increase		3.6 %	6.0 %

There could be different explanations for this success; first the gains in converting non-respondents are better for lower response rates (there may be a level where only marginal gains can be expected from the response rate conversion). However, SLID's test (with only two hundred non-responding households) showed that even with an 88 % response rate, 63 % of the non-respondents were converted one year after. It may be that some non-response is either circumstantial, or that the burden of finishing the LFS and introducing SLID (with no interruption) had a negative impact on response

rate, but that a lag before the next interview could help (it will be interesting to see if that response is maintained longitudinally). Thirdly, extra efforts were put in the June follow-up to explain to the field staff the importance of the operation and this may have impacted on the response rate. Finally, preliminary results seem to indicate that there is more item non-response in the June follow-up. An increase in response rate may be in part due to the fact that interviewers probed to get at least minimal information without the full burden of the interview, which overall improves the response rate but also adds partial response. More studies should be done to substantiate conclusions.

Nonetheless, the results indicate that a gain in response rate can be achieved by follow-up even with a fairly high response rate. The next obvious question is to see if how the conversion of non-respondents impacts on the general response error. The response error has to components, the response rate and the difference in characteristics between respondents and non-respondents. More details can be found in Groves (1989). The benefits of following non-respondents and converting them in respondents has two purposes; first increase the sample size. However, the largest gains happen if there are special characteristics about the converted non-respondents that are different from respondents. In particular, one would like to see if converted non-respondents are closer to "non-converted" respondents than from the respondents. In that case, special adjustments could be made for the converted non-respondents that could compensate for some bias due to non-response. To attempt to measure this second component of the response error, SLID's data were matched to an administrative file.

4. Characteristics of Respondents-Nonrespondents

As it was seen earlier, after the preliminary interview, two options could be envisaged; only respondents could be followed-up or the survey could attempt to do a follow-up of both respondents and non-respondents. The decision was to follow up only respondents because of budget constraints. However, SLID was given the permission to follow-up a small sub-sample of non-respondents. Two hundred non-refusal households were interviewed in the next round for data quality evaluation. 63 % of the two hundred refusal households responded to the first year of the SLID panel (that corresponds to 228 eligible persons). Studies were limited to variables with enough sample.

Two hypothesis were of interest for the study:

- Are converted non-respondents similar to non converted non-respondents ?
- Are attritors similar to converted non-respondents?

In particular, the first hypothesis tries to see the quality gains (other than from a pure number of respondents) achieved by a non-response follow-up. The second hypothesis wants to compare attritors and converted non-respondents to see if there are longitudinal gains to the follow-up conversion. To study the hypothesis SLID was linked to the tax file. The ignorable response mechanism can be tested since SLID income data is very similar to what is reported on the tax file (items that would not be available from the tax were not analysed and excluded from the calculated total income to make the results comparable). Because of the small sample size in SLID, this study was more an evaluation of the extent to which a linkage to an external source was useful in evaluation of the ignorable response mechanism. The small sample size limits the studies that can be done. Studies were restricted to variables that had at least 30 persons in the groups, but *conclusions are very limited*, and *should be used with caution.*

For the comparisons people were classified in four groups; first they have been divided into refusal or respondent to the preliminary interview (almost all non-respondents to the preliminary interview were refusals). Then each sub-group was divided into respondent or non-respondent based on their response to the first year of collection. Non-respondents in the first year could be refusals, but they could also be unable to trace. Problems with the response codes makes it difficult to sub-divide them.

Table 4 shows the distribution of the four categories of response codes, by age of the respondents. As can be seen, there seems to be differences between the distribution of age groups.

Table 4. Age distribution for SLID's sample.

	Refusals preliminary		Respondents preliminary	
	Non-resp YR1 - Non-converted	Resp.YR1 - Converted	Non-resp YR1 - Attritors	Respondents YR1 - Respondents
16-24	24 %	16 %	35 %	17 %
25-54	41 %	44 %	45 %	57 %
55 +	35 %	39 %	19 %	25 %
n	131	228	3643	27587

Similar comparisons were done with other sample characteristics, some differences could be seen in the distribution of marital status (more married people being respondents to both years and more single being found in the "attritors"). But to have a measure of bias, one is really interested in seeing if the response mechanism is ignorable. Since part of SLID's collected information is related to income, SLID's sample was matched to the file of income tax. For more details see Michaud et al. (1995).

There is no unique identifier asked in SLID that would allow the direct linkage to administrative files. SLID full sample was linked to the tax data

file using a record linkage software developed at Statistics Canada (CANLINK). Variables for the linkage were *name, date of birth, sex, marital status, spouse's name, province and postal code.* The fact that most of these variables had been collected with the labour force survey allowed to do a statistical linkage not only with respondents but also with some non-respondents. Table 5 shows the linkage rates.

Table 5. Percentage of SLID sample linked to tax data.

	Refusal Preliminary interview		Responded preliminary interview	
	NC YR1	Converted YR1	Attritors YR1	Respond. YR1
number people aged 16 +	131	228	3643	27587
matched tax	85 (65 %)	183 (80 %)	2210 (61 %)	23585 (86 %)

The matching rate is lower for attritors and "non-converted" respondents. To assess if the non-matching rate can be associated with people that would likely be non-filers (by opposition to get a non-match because the information to do the linkage is missing or wrong), age, sex and marital status were compared for matches and non-matches. Table 6 shows the distributions by age groups. For attritors in particular, it can be seen that there is a high number of unmatched records among people aged between 16 and 24. Similar tables found a higher number of unmatch in the other age groups for married women. For the respondents to the income interview, a large proportion of the unmatched persons (over 80 %) reported an annual income of less than $10,000.

These numbers seem to suggest than there is probably a fair number of unmatches who are actually non filers. For the refusals to preliminary that further refused to responded in the first year of interview, the quality of the matching information may be lower.

Table 6. Comparison of sample distribution and unmatched distribution by age group.

	Refusals preliminary				Respondents preliminary			
	NC YR1		Converted YR1		Attritors YR1		respondents YR1	
	sample	no match	sample	no match	sample	no match	sample	no match
16-24	24 %	46 %	16 %	24 %	35 %	64 %	17 %	36 %
25-54	41 %	34 %	44 %	18 %	45 %	25 %	57 %	32 %
55 +	35 %	20 %	39 %	58 %	19 %	11 %	25 %	32 %

The comparisons of income sources were limited to variables that had enough people. The studied income categories are CPPQPP (Canada or Quebec pension plan benefits), INCTAX (income tax), INVI (interests and

dividends), OASGIS (old age security, guaranteed income supplement and spouse's allowance) and WAGSAL (wages and salary).

The first tables compared to which extent these variables were reported, for respondents vs non-respondents.

Table 7. Reporting of amounts for different income sources.

	Refusal preliminary		Resp preliminary	
	NC YR1	Converted YR1	Attritors YR1	Resp YR1
CPPQPP	24 %	23 %	19 %	18 %
INCTAX	65 %	74 %	63 %	71 %
INVI	47 %	43 %	36 %	40 %
OASGIS	20 %	22 %	16 %	15 %
WAGSAL	64 %	62 %	61 %	68 %

In table 7, it seems that converted non-respondents would be in general closer to non-respondents, and that both categories of refusals to preliminary are different from respondents. The items that are reported more often are items reported to pension amounts. This is consistent to the previous findings that converted non-respondents were older. The "non-converted" non-respondents reported more income from investments and less income tax paid than the converted refusals.

However, because of the very different matching rates between all categories, table 7 may not give a clear picture. In particular, since the previous assumption was that most of the non-matches were probably non-filers, a sensitivity analysis was done, using the extreme assumption that all non-matches were non filers. This assumes that they would have not reported any of the following income sources. The results are presented in table 8.

Results from this table give mixed signals and the similarities from table 7 somewhat vanished.

These numbers seem to suggest than there is probably a fair number of unmatches who are actually non filers. For the refusals to preliminary that further refused to responded in the first year of interview, the quality of the matching information may be lower.

Table 8. Reporting of different income sources, assuming all non matches to tax data are people with no income.

	Refusal preliminary		Respondents preliminary.	
	NC YR1	Converted YR1	Attritors YR1	Resp. YR1
CPPQPP	15 %	18 %	12 %	16 %
INCTAX	42 %	59 %	38 %	61 %
INVI	31 %	34 %	22 %	35 %
OASGIS	17 %	18 %	10 %	13 %
WAGSAL	42 %	50 %	37 %	58 %

The study also looked at the distribution of total income. Results are presented in table 9. For the people linked to the tax file, the median total income of attritors is the lowest. Assuming again that all non filers were low income, this would suggest that the sample will underestimate people in that group and non-converted non-respondents even more.

Table 9. Median of total income reported on tax.

	Refusal preliminary		Respondents preliminary	
	NC YR1	Converted YR1	Attritors YR1	Resp YR1
total income	$ 18412	$ 19352	$ 17747	$ 20109

Although converted respondents resemble more non-converted non-respondents than respondents, the results suggest that converted and non-converted non-respondents may still be different. Having more information on non-respondents (in this case to link them to pertinent extra sources) and a bigger sample may be valuable in helping to assess if estimates are improved by having a non-response follow-up.

5. Further Studies and Conclusions

Other studies have been done on non-respondents for the longitudinal surveys to enhance our understanding of the process. A study has been done with SLID to look at an "ethnic origin" effect. Because of the regional differences in response rates, cultural differences as well as language barriers were pointed out as potential reasons for lower response rate. SLID data was used to verify this hypothesis. At the same time, impact of the topic for the survey and the interviewers themselves were analysed. This is possible since SLID collects two different kind of information yearly, one interview asks labour questions while the second one collects income, and the respondents are contacted (in general) by the same interviewer on both occasion. However, the potential effect of the subject is confounded with the time of collection. Nevertheless, past experience indicated that both collection time has similar effect on response rate, hence it is believed that subject effect would be evaluated properly.

Analyses were done using the first year data of four regional offices out of eight. For each office, a log linear model (Bishop et al. 1980) was fit to the data. Two different sets of models were employed. First, a 2x2x2 table was constructed to analyse the impact of *visible minority* (visible vs non-visible) and *subject* (labour vs income) on *interview response* (total or partial response vs non-response). Second, an *i*x2x2 table was used for studying *interviewers* (*i* different ones) and *subject* on *response*. It was not possible at that time to analyse together all variables because that would

have given too many small cells, and we did not have appropriate software to deal with this issue. Beside, the lack of information on interviewers did not allow us to group them into an objective classification for the analysis.

Table 10 shows the analysis of variance for the *visible×subject×response* model. The variable *response* is always significant at the 1 % level since there are consistently more respondents than non-respondents. The overall response rate per office varies from 75 % to 90 %. Similarly, *visible minority* is significant because more than 95 % of the population are not a member of the visible minority.

Table 10. Maximum-likelihood analysis of variance for visible minority×subject×response model.

		Office A		Office B		Office C		Office D	
Source	Degrees of free-dom	Chi-Square	Prob.[1]	Chi-Square	Prob.	Chi-Square	Prob.	Chi-Square	Prob.
Visible (V)	1	997	0.00	2639	0.00	708	0.00	950	0.00
Response (R)	1	253	0.00	326	0.00	538	0.00	395	0.00
Subject (S)	1	6	0.02	2	0.21	6	0.02	3	0.09
V×R	1	0	0.53	61	0.00	5	0.03	6	0.02
V×S	1	0	0.64	0	0.71	0	0.92	0	0.66
R×S	1	113	0.00	2	0.20	16	0.00	22	0.00
V×R×S	1	2	0.13	14	0.00	4	0.06	2	0.14

[1] Declare significant if probability is less than or equal to 0.01.

To see if a variable has an impact on the *response rate*, one has to look at the interaction terms. For all regional offices, the *subject* has an significant impact on *response rates* at the 1 % level (R×S term). Income seems to be a more sensitive topic than labour. In all but one office, *response rate* for income tends to be smaller than labour. On the opposite, *visible minority* is significant only for office B. For this office, the order 3 term (V×R×S) is also significant. This is cause by the large difference between the members of visible minority and the other people response rate for the labour interview (53 % versus 76 %), while that difference is smaller (69 % versus 76 %) for the income one. Overall for office B, there is no difference between labour and income response rate (around 75 %).

The other model was applied only to the non-visible minority respondents to avoid dealing with small cells. One can see from table 11 that for all offices, the interviewers have a significant effect (I×R) at the 1 % level. Interviewer response rates vary from 70 % to 90 %. For offices A and B the order three term (I×R×S) is also significant, indicating that some interviewers have better results for labour interview and some other have more facilities with income ones. The I×S term is sometime significant because the sample is a little smaller in May (income interview).

Table 11. Maximum-likelihood analysis of variance for interviewerx subjectxresponse model (restricted to non-visible minority members).

		Office A		Office B		Office C		Office D	
Source	Degrees of freedom	Chi-Square	Prob.	Chi-Square	Prob.	Chi-Square	Prob.	Chi-Square	Prob.
Interviewer (I)	off. A=22 off. B=25 off. C= 4 off. D=10	282	0.00	322	0.00	8	0.08	8	0.08
Response (R)	1	1830	0.00	1253	0.00	271	0.00	270	0.00
Subject (S)	1	23	0.00	11	0.00	1	0.30	1	0.30
IxR	22,25,4,10	152	0.00	329	0.00	16	0.00	16	0.00
IxS	22,25,4,10	42	0.01	85	0.00	0	0.99	0	1.00
RxS	1	16	0.00	4	0.04	2	0.21	2	0.21
IxRxS	22,25,4,10	49	0.00	239	0.00	4	0.22	18	0.05

These analyses suggest that the impact of visible minority on response rate is not as strong as it was believed. On the other hand, the subject and the interviewer contribute a lot to the poor/good response rate. It would be relevant to foster interviewer training.

Non-response is a very important issue, especially for a longitudinal survey. A number of results can be drawn from the different studies.

First the length of the interview probably has an impact on response rate. However, it is not clear what is the causal model in our case; what is due to a response burden, what is due to the selected topic and what is caused by an "interviewer burden" (not having a collection window that is long enough).

Response rates can be improved by increasing the collection window (or in our case by transferring the data to the next cycle). However, if the regular field procedures are not changed, the greatest gains happened after the first follow-up. The June follow-up added significantly to the response rates. More efforts had been put in the field procedures, but it was also a more quiet time for collection of other surveys.

Response rates are often used as a proxy measure for quality. When there is a high response rate, even if the non-response mechanism is not ignorable, the estimates will not be affected significantly. However there is a cost, a response burden issue and an ethical issue into following non-respondents, especially when they indicated that they were not willing to participate. These constraints are important to most statistical agencies, and they can not be neglected.

SLID's follow-up showed that when people participated (at least once), a percentage of refusals can be converted. When a high response rate is achieved, the burden of going in a conversion process should be weighted carefully with the gains. SLID attempted to measure the gains of doing a conversion by linking its full sample (of respondents and non-respondents)

to a administrative file. Even if the results are very limited because of the small sample of non-respondents that were followed-up, the exercise seems to be worthwhile. Initial feeling, based on looking at income distribution of total income, is that the converted non-respondents may be different from the other non-respondents. The median income distribution is lower, and there are a few very high income. Attritors also seem different, and the factors that makes a person respondents at the start may be different than the ones that keep a respondent motivated to participate in a longitudinal survey.

As a general conclusion, more information should be collected on non-respondents; whether it is "soft" information such as interviewer's perceptions of the respondent's reasons, more direct information from the respondent (reasons for refusals). Indirect information that allows linkage is valuable to collect. It will usually not be possible to obtain it from all non-respondents but even partial linkage may help to understand the non-response mechanism.

This extra information should have two uses; it could be incorporated in the weighting models to provide better adjustments. It could also be put in a general cost model (see Groves 1989) to try to distinguish when it is worth to follow-up and who should be followed-up.

More research should also be done on contact mechanisms, and on respondent relation material that should be provided and their effects on response and motivation.

References

Binder, D., Michaud, S. and Poirier, C. (1994). Model-based reweighting for non-response adjustment. Turning information into Information Systems. *IRS compilation of selected 1994 ASA papers*, 195–201.

Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1980). *Discrete Multivariate Analysis, Theory and Practice*. MIT Press, Cambridge Massachusetts, U.S.A.

Groves, R. (1989). *Survey Errors and Survey Costs*. John Wiley

Michaud, S., Dolson, D., Adams, D. and Renaud, M. (1995). Combining Administrative and Survey data to reduce respondent burden in longitudinal surveys. *Proceedings ASA, Survey Research Methods*.

Statistics Canada (1995). Response rates for the Longitudinal Social Surveys. Internal Document, Social Survey Methods Division.

SHOULD WE ACCOUNT FOR MISSING DATA IN DUAL SYSTEM ESTIMATION?

Rita J. Petroni, Anne T. Kearney, Machell Kindred Town and Rajendra P. Singh
U.S. Bureau of the Census, USA

The U.S. Census Bureau conducted a Post Enumeration Survey (PES) to evaluate coverage after the 1990 census. To accomplish this, the Bureau selected a sample of census blocks and conducted an independent canvas. Persons and households listed in the census were identified as the E-sample, while those listed in the independent canvas were identified as the P-sample. Analysts matched cases from the two samples and used results to obtain Dual System Estimation (DSE) population estimates. For both samples, statisticians imputed missing data items and used hierarchical logistic regression models to impute unresolved enumeration or match status for persons. For the P-sample, statisticians also adjusted weights to account for noninterviewed households. From the point-of-view of reducting data processing time and effort, an attractive alternative is to treat persons in noninterviewed households, persons with any missing data items, and persons with unresolved enumeration or match status as not captured. That is, ignore such persons by doing no noninterview adjustment, no imputation, and no modeling. This paper analyzes whether this alternative is reasonable for DSE from a statistical viewpoint.

Key words: Noninterview adjustment, imputation, hierarchical logistic regression.

1. Introduction

The U.S. Census Bureau conducted a Post Enumeration Survey (PES) to evaluate coverage after the 1990 census. To accomplish this, the Bureau selected a sample of census blocks and conducted an independent canvas. Persons and households listed in the census from these blocks were identified as the E-sample, while those listed in the independent canvas were identified as the P-sample. Analysts matched cases from the two samples and used results to obtain Dual System Estimates (DSE) for the population. For both samples, statisticians imputed missing data items and used hierarchical logistic regression models to impute unresolved enumeration or

match status for persons. For the P-sample, statisticians also adjusted weights to account for noninterviewed households.

From the point-of-view of reducing data processing time and effort, an attractive alternative is to treat persons in noninterviewed households, persons with any missing data items, or persons with unresolved enumeration or match status as not captured. That is, ignore such persons by doing no noninterview adjustment, no imputation, or no modeling. This paper analyzes the impact on estimates and whether this is reasonable for DSE from a statistical viewpoint. We base our recommendations on comparisons to original DSE estimates since the accepted practice in the statistical milieu is to incorporate noninterview adjustment and imputation into estimation.

After presenting background information on the DSE and compensation for missing data in the 1990 PES in Section II, we provide motivation for the research in Section III, describe the research scope and methodology in Section IV, and provide results and conclusions in Section V.

2. 1990 Post Enumeration Survey Background

Dual System Estimation

We use a statistical model to obtain DSE population estimates from combining the results of the P- and E-samples. The model indicates how errors in the census occur as a stochastic process and how an appropriate estimator is derived. To illustrate, we first present a simple derivation of the DSE estimator using fish in a lake. We follow this by a modification of the estimator to deal with erroneous census enumerations and measurement problems in the PES. (U.S. Bureau of the Census, 1985.)

For this illustration, we want to estimate the number of fish in a lake with no inlets or outlets and where the fish are uniformly and randomly distributed. We use a net to "capture" as many fish as possible in one catch and then count, tag, and release the fish back into the lake. The total fish in this capture is N_1. After the fish have redistributed themselves randomly, but before any births and deaths of fish, we cast the net into the lake again and count the captured fish as well as the tagged or "recaptured" fish. Call the two counts, respectively, N_2 and M. Note that the illustration guarantees the composition of the population remains unchanged over the time of the study. That is, it has no births, deaths, immigrants, or emigrants.

Our objective is to estimate N_T the number of fish in the lake, but we do not know the number of fish not caught in either attempt. To model the capture process, conceptualize the capture of fish as a Bernoulli event. Thus, the counts of fish caught, M, N_1 and N_2, are random variables (the sums of Bernoulli outcomes). Let "a" be the probability a fish is caught in the first

attempt and "b" in the second. Then, if the captures are independent, the expected values of M, N_1 and N_2 are:

$$E(M) = ab(N_T) \quad (2.1)$$
$$E(N_1) = a(N_T) \quad (2.2)$$
$$E(N_2) = b(N_T) \quad (2.3)$$

By substituting the observed values for the expected values and by substituting equations (2.2) and (2.3) into equation (2.1), we obtain:

$$\hat{M} = (N_1)(N_2)/(N_T) \text{ or}$$
$$\hat{N}_T = (N_1)(N_2)/M$$

In a census evaluation application, the theory is exactly the same. Suppose the first capture is the census and the second is the PES. Let N_C be the census count and $\hat{N}_P$ be the weighted sample total from the PES. $\hat{M}$ represents the estimated number of people "captured " in the PES who were enumerated in (i.e. matched to) the census. Then,

$$\hat{N}_T = (\hat{N}_P)(N_C)/\hat{M} \quad (2.4)$$

We adjust N_C in (2.4) to account for imperfections such as erroneous enumerations.

Compensation for Missing Data

The Bureau designed PES operations to obtain interviews for every household in sample, to obtain enough information to resolve the enumeration status of individuals (i.e. to determine whether P-sample individuals were included in the census and whether E-sample individuals were correctly enumerated), and to classify individuals into poststrata defined by demographic characteristics to obtain undercount estimates. Inevitably, however, PES interviewers were unable to obtain interviews for every household. Additionally, ambiguous, conflicting, or missing data from interviewed households made it impossible for clerks to determine with certainty whether some individuals were included in the census. Missing characteristic data were also problematic because we estimate undercount rates for subgroups of the population defined by geography and demographic characteristics (i.e. age, race, sex, and owner/renter status).

This section briefly describes the statistical techniques the Bureau developed for handling noninterviewed households, missing characteristic data, and unresolved enumeration status and the amounts of each. Detailed descriptions are found in Diffendal and Belin (1991).

Noninterviewed households For noninterviewed households in the P-sample, we performed a weighting adjustment within cells defined by sample block cluster and type of structure. The type of structure categories were: (1) one family detached house, (2) one family attached house, (3) building with two or more apartments, and (4) mobile home or trailer, boat, tent, van, etc., and other. The weighting adjustment preserved the sum of the weights within an adjustment cell. 1.6 % of P-sample households were treated as noninterviewed.

Missing characteristic data The Bureau implemented a hot-deck imputation procedure to fill-in a single value for each missing data item for individual characteristics. The hot-deck procedure used certain information about other household members when it was available. When the information was unavailable, the hot-deck procedure imputed values based on a larger reference group. For example, the procedure imputed missing race from other household members when race was observed for the household members; when race was not observed for any household member, it was imputed from persons in nearby housing units. The percentages of individuals with missing characteristics for the P- and E-samples are shown in table 1.

Unresolved enumeration status We developed separate hierarchical logistic regression models for the P- and E-samples to handle unresolved individuals in interviewed households. The models conditioned on available information and used resolved cases. From the fitted models we derived predictions of the probability of match for unresolved P-sample cases and of the probability of correct enumeration for unresolved E-sample cases. In the models, we assumed some effects to be common to all indiviudals and some to vary among defined groups. Covariates in the model included geographical location of the housing unit, demographic characteristics of the individual, predominant racial or ethnic composition of the block, number of individuals in the household, and census or PES processing characteristics. We conducted missing characteristic imputation prior to fitting these models, so that when covariates were missing, their imputed values were used to fit the model. We also embedded other models to assist in predicting probabilities of having been enumerated for unresolved individuals. 1.9 % of the P-sample had unresolved match status, while 1.1 % of the E-sample had unresolved correct enumeration status.

3. Motivation for Research

We can define a continuum for what constitutes "capture by the census" and "capture by PES". At one end we can be very loose about what constitutes capture – accepting late and proxy data, imputing characteristics, etc. At the other end, we can be very strict, only accepting as captures those persons with close to full information who were captured by a given early date. In this case, we consider persons who do not provide this full information as not captured by the census or the PES. We can implement DSE wherever we are on this continuum. To date, the Bureau has defined capture toward the loose end. Using a stricter definition of what constitutes a capture reduces data processing time and effort and may reduce errors in the determination of matches and correct enumerations. It will reduce the census and PES marginal totals and matches so that there are more unobserved persons to be estimated. Going to the stricter end of the continuum may reduce some biases but increase variances and, possibly, correlation bias. (Bell 1995.)

4. Scope of Research and Methodology

This paper reports preliminary research results for two stricter definitions of "capture". We compare national-level estimates produced from these definitions to original 1990 DSE estimates. Based on the outcome of this research, we will determine whether to (1) pursue the use of a stricter definition of "capture" or (2) combine the definitions.

The two stricter definitions are:

- treat persons in noninterviewed households as not captured;
- treat persons with missing characteristics as not captured.

Under the first alternative definition, we do no household noninterview adjustment for the P-sample, but do impute missing characteristics and unresolved enumeration or match status for the P- and E-samples. Under the second alternative definition, we do household noninterview adjustment for the P-sample and impute unresolved enumeration or match status for the P- and E-samples, but exclude persons with any missing characteristics from the P- and E-samples.

Belin et al. (1993) showed that the models for predicting unresolved match and enumeration status have desireable theoretical properties and performed well in practice. Hence, we will not pursue treating cases with unresolved match or enumeration status as not captured.

5. Results and Conclusions

We calculated 357 national-level estimates using the original DSE definition of capture and the two alternative capture definitions along with percent differences between the alternative estimates and the original estimates. Tables 2A and 2B display the percent of alternative estimates which are higher than the original estimates and distributions of percentage differences in original and alternative estimates. Table 3 displays national level estimates for region, tenure, race/Hispanic origin, sex, age, and age/sex and percentage differences in alternative and original estimates. At this point we have not conducted statistical testing, so results are preliminary.

From table 2A, we note the following preliminary points for comparisons of the original DSE estimates to the first alternative definition (treating persons in noninterviewed households as not captured) estimates.

- While a fairly high percent of alternative 1 estimates differ by one percent or less from the original PES estimates, a significant percent differ by more than two percent.
- For all groups except sex, the alternative 1 subgroup estimates differ in closeness to the original estimates.
- For all groups except tenure, sex, and age, the subgroups differ in percent of estimates that are higher than the original estimates.

From table 2B, we note the following preliminary points for comparisons of the original DSE estimates to alternative definition two (treating persons with missing characteristics as not captured) estimates.

- A significant percent of alternative 2 estimates differ by more than two percent from the original estimates.
- For all groups except region, the alternative 2 subgroup estimates differ in closeness to the original estimates.
- For all subgroups, except American Indian on Reservations, most estimates are higher.
- For all groups except region and race/Hispanic origin, the subgroups have roughly the same percent of estimates that are higher than the original estimates.

From table 3, we note the following preliminary points.

- Most alternative 1 estimates are slightly numerically smaller than the original estimates. All except American Indians on Reservations are within one percent of the original estimates.
- Most alternative 2 estimates are numerically higher than the original estimates. All are within two percent of the original estimates.

We need to compute variances and conduct statistical testing before drawing final conclusions regarding the differences noted above. However, preliminarily, we recommend to not drop noninterview adjustment or all persons with any imputed data since a large percent of estimates differ by two percent or more from the original DSE estimates and because of differential effects on subgroups compared to the original estimates.

We are also researching a less strict version of alternative definition 2 to use more of the collected data. With this version we plan to treat only those persons with "most" characteristics missing as not captured.

Acknowledgments: The authors are grateful to Rameswar Chakrabarty for reviewing our manuscript and to Lisa Mundy for secretarial support.

References

Bell, W. R. (1995). Treating Cases as Missed. *From an E-mail message from Bell to Petroni*, March 17, 1995.

Belin, T. R., Diffendal, G. J., Mack, S., Rubin, D. B., Schaefer, J. L. and Zaslovsky, A. M. (1993). Hierarchical Logistic Regression Models for Imputation of Unresolved Enumeration Status in Undercount Estimation. *Journal of the American Statistical Association*, 88, 1149–1159.

Diffendal, G. and Belin, T. (1991). *Results of Procedures for Handling Noninterviews, Missing Characteristic Data, and Unresolved Enumeration Status in 1990 Census/Post-Enumeration Survey*. STSD Decennial Census Memorandum Series #V-112, July 1, 1991, Bureau of the Census.

U.S. Bureau of the Census (1985). Evaluating Censuses of Population and Housing, *Statistical Training Document. ISP-TR-5, Chapter 2*, Washington, D.C.

Table 1. Percent of Persons with Missing Characteristics.

CHARACTERISTICS	P-SAMPLE	E-SAMPLE
Age	0.725	2.582
Race	0.935	3.498
Sex	0.532	1.185
Hispanic Origin	2.293	10.553
Tenure (Owner/Renter)	2.389	3.049
Structure (Type of Building)	1.153	2.515
Marital Status	1.610	2.543
Relationship to Reference Person	0.042	2.344

Table 2A. Comparison of Alternative One Estimates to Original DSE Estimates.

Characteristics	Number of Estimates	% of Estimates Higher[1]	%[2] of Estimates Differing from Original DSE Estimates by: <=1%	±(1.01,2)%	±(2.01,5)%	±(5.01,20)%
REGION						
Northeast	70	51	64	10	19	6
South	70	31	59	20	19	1
Midwest	70	56	70	15	12	2
West	70	46	49	26	17	8
TENURE						
Owner	175	50	75	14	8	4
Renter	175	47	46	25	23	9
RACE/ORIGIN						
Non.Hisp.White, etc.[3]	168	39	73	15	11	1
Black	84	60	51	23	16	9
NonBlack Hisp.	84	55	42	28	18	11
Asian or P.I.	14	50	65	0	21	14
AI on Reserv.	7	71	28	0	42	29
SEX						
Male, 18+	153	51	60	15	15	9
Female, 18+	153	44	57	20	16	7
AGE						
0-17	51	55	62	24	14	0
18-29	102	50	51	21	19	10
30-49	102	49	58	20	17	5
50+	102	44	66	15	11	8
AGE/SEX						
0-17	51	55	62	24	14	0
18-29, Male	51	53	61	16	16	8
18-29, Female	51	47	42	26	22	12
30-49, Male	51	53	58	16	20	6
30-49, Female	51	45	59	24	14	4
50+, Male	51	47	60	16	12	12
50+, Female	51	41	72	14	10	4

1 Significance testing has not yet been conducted.
2 Due to rounding % may not add to 100.
3 Non-Hispanic White, Other, American Indian not on reservation.

Table 2B. Comparison of Alternative One Estimates to Original DSE Estimates.

Characteristics	Number of Estimates	% of Estimates Higher[1]	%[2] of Estimates Differing from Original DSE Estimates by: <=1%	±(1.01,2)%	±(2.01,5)%	±(5.01,20)%
REGION						
Northeast	70	76	48	22	22	6
South	70	83	50	27	15	7
Midwest	70	91	45	22	27	7
West	70	83	49	27	16	7
TENURE						
Owner	175	80	57	23	17	3
Renter	175	81	29	27	33	12
RACE/ORIGIN						
Non.Hisp.White, etc.[3]	168	84	60	26	12	2
Black	84	86	22	28	33	15
NonBlack Hisp.	84	67	30	20	38	10
Asian or P.I.	14	93	29	21	50	0
AI on Reserv.	7	29	58	43	0	0
SEX						
Male, 18+	153	81	34	28	27	11
Female, 18+	153	79	44	25	26	5
AGE						
0-17	51	76	67	18	14	2
18-29	102	80	32	31	31	7
30-49	102	83	39	35	21	5
50+	102	76	47	14	28	12
AGE/SEX						
0-17	51	76	67	18	14	2
18-29, Male	51	80	30	26	36	10
18-29, Female	51	80	34	37	26	4
30-49, Male	51	86	30	43	20	8
30-49, Female	51	80	49	28	22	2
50+, Male	51	76	43	16	26	16
50+, Female	51	76	51	12	30	8

1 Significance testing has not yet been conducted.
2 Due to rounding % may not add to 100.
3 Non-Hispanic White, Other, American Indian not on reservation.

Table 3. Percent Differences in Original PES and Alternative Total Person Estimates.

	Original	Alternative 1	% Diff.	Alternative 2	% Diff.
REGION					
Northeast	48,141,043	48,183,644	0.09	48,561,945	0.87
South	74,278,336	73,988,436	-0.39	75,014,896	0.99
Midwest	56,855,276	56,800,383	-0.10	57,255,503	0.70
West	45,572,499	45,491,733	-0.18	46,026,297	1.00
TENURE					
Owner	162,068,429	161,977,769	-0.06	162,956,149	0.55
Renter	85,844,450	85,564,956	-0.33	87,282,882	1.68
RACE/ORIGIN					
NonHispWhite, etc[1]	188,055,098	187,543,669	-0.27	189,433,697	0.73
Black	30,395,440	30,419,744	0.08	31,003,711	2.00
NonBlack Hisp	22,060,514	22,169,784	0.50	22,280,575	1.00
Asian or Pac Isl.	7,401,827	7,409,529	0.10	7,521,048	1.61
Amer.Ind. on reserv	425,109	437,312	2.87	421,938	-0.75
SEX					
Male 18+	87,501,561	87,421,928	-0.09	88,654,355	1.32
Female 18+	95,376,676	95,165,189	-0.22	96,257,797	0.92
AGE					
0-17	65,459,751	65,392,920	-0.10	65,748,817	0.44
18-29	48,284,368	48,199,785	-0.18	48,999,977	1.48
30-49	73,408,700	73,240,304	-0.23	74,301,422	1.22
50+	61,185,169	61,147,028	-0.06	61,610,753	0.70
AGE/SEX					
0-17	65,459,751	65,392,920	-0.10	65,748,817	0.44
18-29, Male	24,019,892	24,016,075	-0.02	24,450,267	1.79
18-29, Female	24,264,476	24,183,710	-0.33	24,549,710	1.18
30-49, Male	36,208,596	36,117,685	-0.25	36,724,847	1.43
30-49, Female	37,200,104	37,122,619	-0.21	37,576,574	1.01
50+, Male	27,273,073	27,288,168	0.06	27,479,241	0.76
50+, Female	33,912,096	33,858,860	-0.16	34,131,512	0.65

[1]Non-Hispanic White, Other, American Indian not on reservation.

PREVALENCE OF ANSWERING MACHINE USAGE IN AGRICULTURAL SURVEY POPULATIONS

Jaki S. Stanley and Terry P. O'Connor
National Agricultural Statistics Service, U. S. Department of Agriculture

Answering machine usage was tracked for a large CATI Agricultural Survey. Whether an attempted contact ever reached an answering machines and the final disposition of CATI sample unit was recorded for 4 successive quarters of data collection. Respondents do not appear to use answering machines to screen calls; the proportion of refusals among answering machine contacts was the same as those respondents where answering machines were never contacted. The prevalence of answering machines varies by state. However, at the present time, there appears to be little adverse impact on response rates in agricultural surveys due to answering machines.

Key words: CATI, response rate, refusals.

1. Introduction

The National Agricultural Statistics Service makes over 750,000 survey data collection contacts each year in order to collect information about the nation's agriculture. Over 70 % of these contacts are made by telephone. Previous studies have shown a prevalence of approximately 25 % of U.S. households with answering machines (Tuckel and Feinberg 1991, Oldendick and Link 1994). However, with the price of answering machines decreasing, the prevalence of answering machines is most likely increasing. The increasing proliferation of answering machines has the potential to severely affect data that are collected by telephone.

Answering machines also have the potential to affect survey indications by altering the representativeness of the survey sample contacted if respondents using answering machines are significantly different from those who do not. Oldendick and Link (1994) found that characteristics of answering machine owners contacted in random digit dialed household surveys in South Carolina varied over income level, education, age, and race. However, none of the demographic characteristics varied with respect to reported call screening. Therefore, it appeared that sample representativeness was not jeopardized by respondents differentially using call screening to remain inaccessible.

To date, answering machine incidence has been studied in localized or national household samples, but very little has been done to examine its impact specifically in rural agricultural populations. Studies of national household samples have found that answering machine usage is significantly lower in rural households than in urban households (Oldendick 1993, Oldendick and Link 1994). Rural households also reported significantly less usage of answering machines for call screening.

This paper discusses the prevalence and impact of answering machines in agricultural survey populations.

2. Data and Methods

The data discussed here were collected in the NASS Agricultural Survey (AS). This survey is conducted in every state except Alaska and Hawaii in June, September, December and March of each year. Information is collected about crops, stocks and livestock from list frame samples of over 70,000 respondents each quarter. Data collection in NASS is decentralized with each of 44 State Statistical Offices (SSOs) collecting information primarily from respondents operating in their state. Data collection periods are quite short, typically only two weeks long.

Approximately 35 % of this sample is typically contacted for computer assisted telephone interviews (CATI), approximately 25 % contacted by non-CATI telephone interviews, 15 % contacted by face to face interviews, 3 % contacted by mail, with the rest making up estimated data, refusals and inaccessibles (Ramirez and Tesky 1995).

The data shown here were collected only for CATI interviews, in all states except for Colorado, Indiana, New Jersey, Wyoming, Alaska and Hawaii (these states either did not complete any CATI interviews or used different CATI software which did not collect answering machine information). Shown below is the disposition of the national sample for 4 quarters of the AS. The first line is for contacts where an answering machine was NEVER contacted, the second line shows contacts where an answering machine was contacted at least one time. This table indicates the prevalence of answering machine contacts, NOT the prevalence of answering machines owned. If an answering machine was owned by the respondent, but not in use at the time of the call, this is not recorded.

Table 1. Number of Answering Machine Contacts by Quarter.

	Dec 1994	March 1995	June 1995	September 1995	TOTAL
No Answering Machine Contacted	29,522 86.4 %	29,319 84.3 %	28,875 79.9 %	26,620 80.31 %	114,336 82.73 %
Answering Machine Contacted At Least Once	4,635 13.6 %	5442 15.6 %	7268 20.1 %	6524 19.68 %	23,869 17.27 %

The overall percent of contacts with answering machines is lower than those cited for full national household samples, but is consistent for figures cited for rural populations (Oldendick 1993).

3. Answering Machine Usage and Response Rates

While the number of answering machines in use in agricultural populations can be fairly high, its impact on response rates is still fairly minimal, with many answering machine contacts ultimately resulting in completed interviews. Shown below is the prevalence of answering machine contacts broken down by final outcome of the interview (cases for which data was estimated are not included). The first column shows the number of completed responses, the second column shows the number of people who were contacted and refused to provide information and the last column shows the number of people who were never contacted and provided no information.

Table 2. Call Outcomes for Answering Machine/No Answering Machine Contacts by Quarter.

		Completed Interview	Refusal	Inaccessible
December 1994	No Answering Machine Contacted	24823 88.67%	3814 87.68%	857 72.8%
	Answering Machine Contacted At Least Once	3171 11.33%	536 12.32%	320 27.19%
March 1995	No Answering Machine Contacted	24892 85.02%	3596 84.45%	831 67.84%
	Answering Machine Contacted At Least Once	4386 14.98%	662 15.55%	394 32.16%
June 1995	No Answering Machine Contacted	24198 80.74%	3744 79.49%	933 63.64%
	Answering Machine Contacted At Least Once	5771 19.26%	966 20.51%	533 36.36%
September 1995	No Answering Machine Contacted	22460 80.92%	3458 81.17%	702 62.18%
	Answering Machine Contacted At Least Once	5295 19.08%	802 18.83%	427 37.82%
All Quarters	No Answering Machine Contacted	96373 83.81%	14612 83.13%	3323 66.50%
	Answering Machine Contacted At Least One	18623 16.19%	2966 16.87%	1674 33.50%

As is shown in Table 2, many of the respondents who have answering machines do ultimately complete interviews. In addition, the proportion of answering machine contacts that ultimately refused is similar to the proportions for those completing interviews. This suggests that answering machines are not being used to screen unwanted calls. Instead, respondents with answering machines are willing to answer the phone and provide an

explicit refusal. If answering machines were being used to screen calls, the proportion of refusals with answering machines would be much lower. Instead these people would become inaccessibles.

This implies that the higher percentage of inaccessibles that have answering machines are simply unavailable, and for this population answering machines are not being widely used by "cocoon-ers" to screen calls. Instead, they appear to be used more by "connectors" who wish to be able to receive messages while they are unavailable.

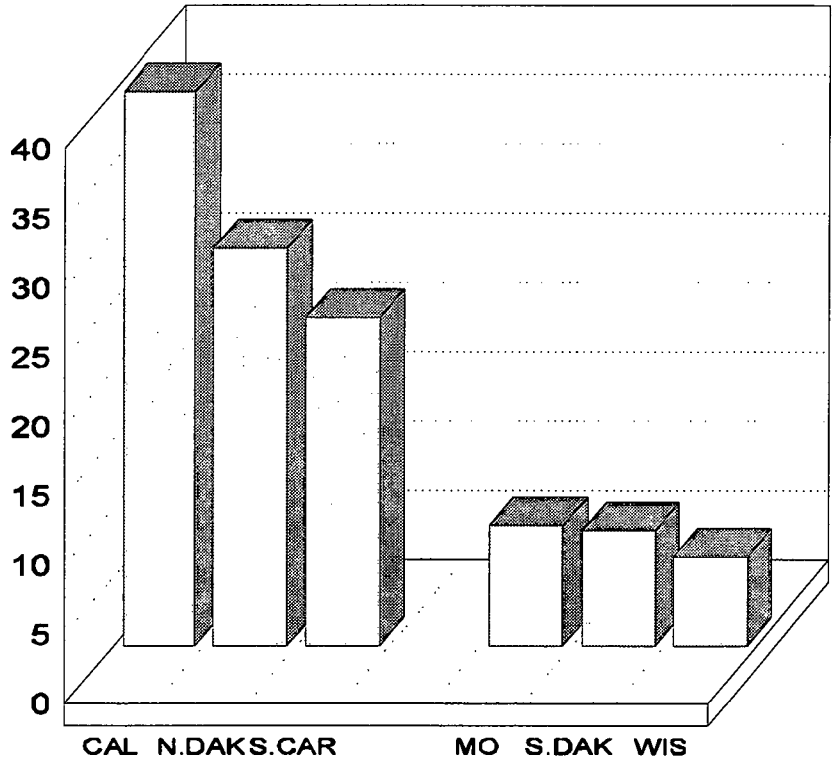
As was seen in Table 1, the proportion of answering machines reached in December is noticeably lower than the proportion reached in June. Agricultural production is a very seasonal activity, this is reflected in the proportion of answering machines encountered. Fewer answering machines are contacted in December when agricultural producers are spending less time working outside, compared with parts of the year when more activity (planting, harvesting, etc.) may be taking place.

4. Answering Machine Usage and Geographic Location

For the agricultural survey population, many things vary depending on the location of the respondent. The types of agricultural commodities produced, the extent of urbanization, the local climate, and other things vary from state to state. The prevalence of answering machines also appears to vary by location.

At the present time, NASS does not have a standard national policy for handling answering machine contacts. Each SSO is free to set its own policy and approach to dealing with these contacts. At the present time, most SSOs treat answering machines the same way as contacts where there is no answer and do not leave a message. Some SSOs have adopted the policy of leaving a tollfree office telephone number and asking the respondent to return the call. There is no data tracking the success of this approach, although SSO personnel report that they do receive some responses this way. However, SSOs who report leaving messages on answering machines do not appear to have higher response rates than states that do not.

Fig. 1. Percent of Answering Machine Contacts, Top and Bottom Three States (All Quarters).
(Notation: California, North Dakota, South Carolina, Montana, South Dakota, Wisconsin.)



5. Discussion

The most obvious affect on survey data collection is through depressed response rates. Answering machines may potentially be used to screen calls by respondents choosing not to answer calls from survey interviewers. This will obviously contribute to a decrease in response rates for telephone surveys. At the present time, our data indicate that answering machines are not a significant problem for NASS.

Answering machines do, however, also hold the potential to increase response rates. The answering machine is an additional opportunity to give information to the respondent to motivate their participation. This opportunity is not available when a number is reached for which the telephone rings but is never answered. Tuckel and Feinberg (1991) found that in households where an answering machine was the first contact, response was more likely than in households where the first contact resulted in either no answer or a busy signal.

Piazza (1993) found that calls made to telephone numbers that had previously had an answering machine were more likely to result in a completed interview at certain times of the week. The times of the week resulting in the highest number of post answering machine completions were Saturday morning from 9 a.m.–12 noon and Sunday through Thursday evenings from 6 p.m. to 10 p.m.

While studies of national household samples are informative for rural samples, the differing characteristics of these samples most likely lead to different conclusions about the best strategies for reaching these respondents. For example, most household surveys generally target evening hours for calling in order to reach respondents after they have returned home from their place of employment. However, for rural (specifically, farm and ranch operators) populations the optimum times for calling can be quite different. NASS typically calls respondents in the evenings also, but can also have high rates of contact during early morning hours and at midday. These are times not typically called for household surveys. Indeed, in his study, Piazza (1993) began calling at 9 am, whereas calls as early as 6 or 7 am are not unheard of for (and may be requested by) farmers and ranchers.

Each SSO has discretion over how to handle answering machine contacts. The data indicate that current SSO practices are effective in maximizing completed interviews and minimizing refusals and inaccessibles when an answering machine is contacted at least once during an AS. The next stage for this research may be to investigate the effectiveness of specific SSO practices.

References

Oldendick, R. W. (1993). The Effect of Answering Machines on the Representativeness of Samples in Telephone Surveys. *Journal of Official Statistics*, 9(3), 663–672.

Oldendick, R. W. and Link, M. W. (1994). The Answering Machine Generation: Who Are They and what Problem Do They Pose for Survey Research. *Public Opinion Quarterly*, 58, 264–273.

Piazza, T. (1993). Meeting the Challenge of Answering Machines. *Public Opinion Quarterly*, 57, 219–231.

Ramirez, J. and Tesky, M. (1995). 1992/1993 *Agricultural Surveys, Administration Analysis*. NASS Staff Report SAB Number 95–01, US Department of Agriculture.

Tuckel, P. S. and Feinberg, B. M. (1991). The Answering Machine Poses Many Questions for Telephone Survey Researchers. *Public Opinion Quarterly*, 55, 200–217.

EVENT HISTORY ANALYSIS OF RESPONSE TIMES: TESTING EFFECTS OF INFORMED CONSENT IN THE DUTCH NATIONAL TRAVEL SURVEY

Jan A. van den Brakel, Martin Luppes and Ger Moritz
Statistics Netherlands

An embedded field experiment was conducted in order to determine the influence of an informed consent letter on response behaviour and estimates of population parameters of the Dutch National Travel Survey (NTS). Event history analysis and logistic regression analysis were applied to test the hypothesis that the informed consent letter has no effect on response times respectively response rates of the NTS. No significant differences in response times, response rates, and estimates of population parameters between the experimental group and the control group could be found.

Key words: Event history analysis, informed consent, logistic regression, response rates, response times.

1. Introduction

Increasing nonresponse rates to surveys is one of the problems a national statistical office is faced with. In order to improve response rates, much research has been done on different factors influencing response behaviour. One of these factors is the contents of advance letters needed for informed consent. However, the contents of advance letters is not a clear or decisive factor in improving response rates (and/or data quality). The effect of different advance letters on response behaviour and data quality in the NTS is studied by means of an embedded field experiment.

Logistic regression models are often used to analyse response rates. However, these models only use the binary information of response or nonresponse. If interest is also focused on the duration time until a completed questionnaire is returned (in this paper called response time), event history models may be more appropriate. In this article the outcomes of the embedded field experiment are analysed by means of such models.

In section 2 the NTS is described. The purpose of this study and the experimental design applied are described in section 3. The use of logistic regression analysis and event history analysis for response rates and

response times are described in section 4. In section 5 the results of the experiment are presented. Finally section 6 contains the conclusions of this study.

2. The National Travel Survey

The Dutch National Travel Survey (NTS), conducted by the department of Traffic and Transport of Statistics Netherlands, started in January 1978. The NTS describes the journeys and patterns of mobility of the Dutch population.

The target population of the NTS is the resident population of the Netherlands. The sample frame is the so-called Geographic Basic Register, a file of postal addresses from which business addresses have been eliminated. The sample is based on a stratified two-stage sample of addresses with Province and Urbanisation as stratification variables. The primary sampling units are municipalities, and the secondary sampling units are addresses. The sample is self-weighted, i.e. the inclusion probabilities of the addresses are equal. All households living on the selected addresses are included in the sample, and are spread randomly over all the days of the year. The date thus allotted to each household is the date for which the respondents will be requested to complete a journey diary.

The data are collected in a telephone interview as well as a journey diary sent by mail. The survey is announced in an advance letter. A few days after sending this letter, one of the members of the household is contacted by telephone and asked to provide some information about the household situation. Next, diaries are mailed to all household members. Each individual is asked to keep record of all of his/her journeys for one day. There are at most two recalls, where new diaries are sent to the respondent when he/she fails to return a completed diary on time. Again, the respondent is asked to keep record of all of his/her journeys on a new predetermined interview day (respectively seven and fourteen days after the first interview day). If the respondent does not return a completed diary after the second recall, he or she is considered a nonresponse.

3. Experimental design

In the standard advance letter of the NTS, respondents are informed about the goal of the survey, the voluntariness of participation and the guarantee of confidentiality and anonymity of Statistics Netherlands. Participation in the NTS is based on passive consent, i.e. no explicit consent is asked, either by the interviewer or by signature. An important element of informed consent in voluntary surveys is the guarantee of confidentiality and anonymity.

Because of this, only restricted, anonymous microdata of the NTS are released for statistical purposes.

The Ministry of Transport, Public Works and Water Management (further referred to as the Ministry) requests more detailed information in the NTS. Providing less restricted, but still anonymous microdata to the Ministry is legally possible when the potential respondent is informed that both Statistics Netherlands and the Ministry will conduct the survey and both guarantee confidentiality and anonymity. This implies the inclusion of an informed consent paragraph as a standard in the advance letter of the NTS. Whether or not the informed consent paragraph can be included as a standard in the advance letters of the NTS, depends partly on its effects on response behaviour and estimates of population parameters. Letters containing this paragraph should not result in significant changes in both response behaviour and estimates of the population parameters. To this end an embedded field experiment was conducted in which the advance letter is used as the experimental factor.

Three questions had to be answered within this study:

1 what is the effect of the experimental letter on response rates?
2 does the experimental letter affect response times?
3 is there a difference in the estimates of the population parameters?

Possible differences in response rates were tested by means of logistic regression analysis, possible differences in response times by means of event history analysis, and possible differences in the estimates of population parameters by means of two parameter free tests (the two sample test of Wilcoxon and the two sample test of Smirnov).

The sample size of the field experiment amounted 10 362 households with a telephone number. This sample was randomly divided into two interpenetrating subsamples with sample sizes of 8296 and 2066 households. The households in the largest subsample are assigned to the control group and the households in the smallest subsample to the experimental group. The control group received the standard advance letter in which the promise of confidentiality implied the dissemination of restricted anonymous microdata to the Ministry. The experimental group received the experimental advance letter in which an informed consent paragraph is provided. Apart from the difference in the advance letters, both groups were treated the same. This test was carried out as a blind experiment in order to avoid that interviewers would adjust their behaviour (possibly unconsciously) because they knew that they participated in an experiment.

4. Analysis of Response Rates and Times

Logistic Regression

In the analysis of response behaviour, interest goes out to the effects of explanatory variables on response rates. Therefore a model is required which describes the effects of explanatory variables on response rates. Let y_i denote the dependent binary variable which describes response behaviour for individual i. This dependent variable can take values $y_i = 1$ (response) and $y_i = 0$ (nonresponse). Instead of modeling the effect of explanatory variables directly on the value of the discrete variable y_i, it is more natural to use the probability of $y_i = 1$ as the dependent variable. It holds that $E(y_i) = 1xP(y_i=1) + 0xP(y_i=0) = P(y_i=1)$. Let $E(y_i \mid \mathbf{x_i}) = P(y_i=1 \mid \mathbf{x_i})$ denote the probability of response at the ith setting of values of k explanatory variables $\mathbf{x_i} = (x_{i1}, x_{i2}, ..., x_{ik})^t$ (the superscript t stands for the transposed sign). A commonly used model to describe the probability of response as a function of continuous and/or discrete variables is the logistic regression model (Agresti 1990, Ch.4):

$$p(y_i = 1 | \mathbf{x_i}) = \frac{\exp(\alpha + \mathbf{b}^t \mathbf{x_i})}{1 + \exp(\alpha + \mathbf{b}^t \mathbf{x_i})}, \qquad (4.1)$$

equivalently,

$$\frac{p(y_i = 1 | \mathbf{x_i})}{1 - p(y_i = 1 | \mathbf{x_i})} = \exp(\alpha + \mathbf{b}^t \mathbf{x_i}), \qquad (4.2)$$

with α the intercept parameter and $\mathbf{b} = (\beta_1, \beta_2, ..., \beta_k)^t$, the regression coefficients of the explanatory variables. This model can also be expressed as the so-called logit model (Agresti 1990, Ch.4):

$$\log\left(\frac{p(y_i = 1 | \mathbf{x_i})}{1 - p(y_i = 1 | \mathbf{x_i})}\right) = \alpha + \mathbf{b}^t \mathbf{x_i}, \qquad (4.3)$$

Formula (4.2) provides a basic interpretation for the regression coefficients. The odds of response instead of nonresponse increases multiplicatively by $\exp(\beta_k)$ for every unit increase in x_{ik}, controlling for other covariates.

If logistic regression analysis is applied to response rates, only the binary information of responding or not responding is used. The information on response times is not taken into account. If interest goes out to the duration time of returning completed questionnaires, response times should be included into the analyses. When nonrespondents are considered to have right censored response times, it is recognised that event history analysis is an appropriate way of analysing response times. For an introduction into the theory of event history analysis (also called survival analysis) we refer to Cox and Oakes (1984), Kalbfleisch and Prentice (1980) or Yamaguchi (1991).

In event history analysis, the time until a defined event occurs (in this case the response) is the subject of study. Theoretically the time until an event occurs is continuous. In this study, the event of interest is the completion of a questionnaire by the respondent. The occurrence of such events are only observed at discrete time points, usually a day. In the NTS, respondents are asked to fill in a journey diary on a predetermined interview day. If they refuse to respond, at most two recalls with a request to complete a new diary on a new predetermined interview-day are sent to the respondent (see section 2). So, in this case, there is a discrete time process with time steps equal to the time between each of these predetermined interview days. Note that there is a maximum of three time steps. Two central concepts in event history are the risk set and the hazard rate. The *risk set* is defined as the set of individuals who are at risk of event occurrence at time t_j, given that the event did not occur before time t_j, $j = 1, 2, ..., J$. In discrete time, the *hazard rate* is defined as the (conditional) probability of having the event at time t_j, given that the event did not occur before time t_j, $j = 1, 2, ..., J$. Let T_i denote a discrete random variable that indicates the time of an event for individual i. The event occurring at time t_j is denoted by $T_i = t_j$, $j = 1, 2, ..., J$. The hazard rate at time t_j for the ith individual can be denoted as $\lambda_i(t_j) = P(T_i=t_j \mid T_i \geq t_j)$. Assuming that the hazard rate is the same for all individuals, i.e. $\lambda_i(t_j)=\lambda(t_j)$, estimates of the hazard rate at each time point t_j can be obtained by dividing the number of individuals having the event at time t_j by the number of individuals at risk at time t_j.

Since in discrete time the hazard rate is characterised by conditional probabilities, the logistic regression model can be used to specify how the hazard rate depends on explanatory variables. In this so-called discrete time logit model, the conditional probabilities at each discrete time point t_i are modelled as follows (Yamaguchi 1991):

$$\frac{\lambda(t_j|\mathbf{x_i})}{1-\lambda(t_j|\mathbf{x_i})} = \frac{\lambda_0(t_j)}{1-\lambda_0(t_j)}\exp(\mathbf{b}^t\mathbf{x_i}),\ (j=1,2,\ldots,J), \qquad (4.4)$$

where $\lambda(t_j|\mathbf{x_i}) = P(T=t_j|T\geq t_j;\mathbf{x_i})$ denotes the hazard rate at time point t_j at the ith setting of values of k explanatory variables $\mathbf{x_i} = (x_{i1}, x_{i2}, \ldots, x_{ik})^t$ and $\lambda_0(t_j)$ denotes the so-called baseline hazard. The *baseline hazard* is characterised by the conditional probabilities for cases in which the covariate vector $\mathbf{x_i} = \mathbf{0}$. Clearly the model assumes that for every individual the odds of having the event at each discrete time t_j is proportional to the odds of having the event for some specific individual who represents the set of baseline states of covariates. The interpretation of the regression coefficient is as follows. The odds of having the event at each discrete time point t_j is $\exp(\mathbf{b}^t\mathbf{x_i})$ times higher for individuals characterised by covariates $\mathbf{x_i}$ compared with individuals in the baseline group for all $j = 1, 2, \ldots, J$. An increase in one unit of covariate x_{ik}, while controlling for other covariates, increases the odds of having the event $\exp(\beta_k)$ times. Logistic regression model (4.4) can also be expressed in logit form:

$$\log\left(\frac{\lambda(t_j|\mathbf{x_i})}{1-\lambda(t_j|\mathbf{x_i})}\right) = \alpha_j + \mathbf{b}^t\mathbf{x_i},\ (j=1,2,\ldots,J), \qquad (4.5)$$

with

$$\alpha_j = \log\left(\frac{\lambda_0(t_j)}{1-\lambda_0(t_j)}\right).$$

5. Results

Response Rates and Times

The experiment started with a number of 8296 households reachable by telephone in the control group and 2066 households reachable by telephone in the experimental group. The response results on the telephone interview are summarised in Table 1.

Table 1. Response results for households on the telephone interview.

	response	nonresponse	total
experimental group	1566 (75.8 %)	500 (24.2 %)	2066
control group	6108 (73.6 %)	2188 (26.4 %)	8296
total	7674 (74.1 %)	2688 (25.9 %)	10362

A total of 4018 persons out of the 1566 households willing to participate with the survey in the experimental group, as well a total of 15293 persons out of the 6108 households in the control group received a journey diary. The response results on these diaries are summarised in Table 2.

Table 2. Response results for persons on the journey diaries.

	response	nonresponse	total
experimental group	3131 (77.9 %)	887 (22.1 %)	4018
control group	11913 (77.9 %)	3380 (22.1 %)	15293
total	15044 (77.9 %)	4267 (22.1 %)	19311

In line with the first question stated in section 3 we formulated the first null hypothesis as follows: there is no difference in the response rates between the experimental group and the control group. This hypothesis is tested against the alternative that there is a difference between the response rates. The logistic regression model described in subsection 4.1 is used to test this hypothesis. To this end, the response behaviour of the individuals is modelled as a binary variable in a logistic regression model. A dummy variable indicating that an individual belongs to the experimental group or the control group (the so-called experimental factor) is included in the model as an explanatory variable. The effect of the experimental factor can be estimated and its significance can be tested, controlling for other explanatory variables.

The results of the logistic regression analysis for the telephone interview and the journey diaries are summarised in Table 3. In the logistic regression model for the response of the telephone interview, the experimental factor is the only explanatory variable, because no auxiliary information on the nonrespondents was available. In the logistic regression model for the response of the journey diaries, the stratification variables (Province and Urbanisation) were used as auxiliary variables to control for. The effects of the stratification variables appear to be highly significant at a significance level of 0.01. These results are not presented here because interest is focused on the effects of the experimental factor.

Table 3. Results of logistic regression analysis.

experimental factor	$\hat{\beta}$	$\exp(\hat{\beta})$	$\hat{\sigma}(\hat{\beta})$	p-value (two tailed)
telephone interview	0.1151	1.122	0.0571	0.044
journey diaries	0.0018	1.002	0.0427	0.966

At a significance level of 0.01 no significant difference between the response rates of the experimental group and the control group can be found either for the telephone interview or for the journey diaries. Although the regression coefficient of the experimental factor has a small p-value in the model for

the telephone interview, its influence on the odds of the response (exp(β)) is small.

In line with the second question stated in section 3 we formulated the second null hypothesis as follows: there is no difference in response times between the experimental group and the control group. This hypothesis is tested against the alternative hypothesis that there is a difference between the response times. Focusing on the duration time of responding, event history analysis can be applied to the response times of the journey diaries. Table 4 shows estimates of the hazard rates for the experimental group and the control group. The event of interest is whether a respondent has returned a completed journey diary. Respondents who did not return a diary after two recalls are censored at time point $t_j = 3$. Respondents who returned a diary completed incorrectly are censored at the time of completion.

Table 4. Estimated hazard rates of the journey diaries.

		experimental group			control group		
response time		*response*	number at *risk*	estimated hazard *rate*	*response*	number at *risk*	estimated hazard *rate*
1	(first diary)	2670	4018	0.665	10003	15293	0.654
2	(first recall, second diary)	370	1227	0.302	1478	4871	0.303
3	(second recall, third diary)	91	834	0.109	432	3260	0.132

Differences in response times between the experimental group and the control group show up as differences in the estimated hazard rates for both groups. The hazard rate of the response behaviour can be analysed with the discrete time logit models as described in subsection 4.2. To investigate differences between the hazard rates of the experimental group and the control group (the second hypothesis), a dummy indicating if an individual belongs to the experimental or control group (experimental factor) can be included in the model as an explanatory variable. To control for the stratification variables from the sampling design, the variables Province and Urbanisation are also included in the model as explanatory variables. The estimation results of the discrete time logit model are summarised in Table 5. The effects of the stratification variables are very significant at a significance level of 0.01. The estimation results of these variables are omitted because interest goes out to the effects of the experimental factor.

Table 5. Estimation results of the discrete time logit model of the hazard rate of the journey diaries.

	$\hat{\beta}$	$\exp(\hat{\beta})$	$\hat{\sigma}(\hat{\beta})$	p-value (two tailed)
experimental factor	0.0150	1.015	0.0318	0.638

At a significance level of 0.01, no significant difference between the hazard rates of the experimental group and the control group were found. The estimated regression coefficient of the experimental factor indicates that the odds of having an event for individuals in the experimental group is 1.015 times higher than the odds of having an event for individuals in the control group.

Mobility

In order to investigate if the informed consent letter has an affect on the estimates of three population parameters of the NTS, e.g. by means of selective nonresponse, hypotheses concerning these mobility parameters were tested. The analysis is restricted to the distance travelled per person per day, the kilometre per journey, and the number of journeys per person per day. Some descriptive statistics of the sample distributions of these parameters are summarised in Table 6.

Table 6. Descriptive statistics of the distributions of three population parameters of the NTS for the experimental group and the control group.

	experimental group			control group		
	1	2	3	1	2	3
median	13.00	3.40	3.00	12.50	3.33	3.00
mean	32.06	9.49	3.55	31.69	9.50	3.48
variance	2680.48	319.17	6.47	2898.51	357.59	6.67
skewness	3.70	4.64	0.81	4.03	5.48	0.88

1: distance travelled per person per day
2: kilometre per journey
3: number of journeys per person per day

Because the sample distributions of these parameters are very skew to the right, it is not realistic to assume that these data follow normal distributions. Because the t-test and the F-test are not robust against outliers and violations of normality, two parameter-free tests are used to test the third hypotheses, namely that the experimental advance letter does not affect estimates of the population parameters. The first parameter-free test is the two sample test of Wilcoxon which has a large power for alternative hypotheses concerning differences in location of the sample distributions. The second parameter-free test is the two sample test of Smirnov (also known as the two sample test of Kolmogorov Smirnov). Under the alternative hypothesis, the test of Smirnov has a large asymptotical power for all sorts of possible differences which can exists between two sample distributions. For a detailed description of these tests, see Lehmann (1975). The results of these tests are summarised in Table 7. At a significance level of 0.01, none of the three null hypotheses can be rejected.

Table 7. Analysis results of the two sample test of Wilcoxon.

	p-value (2-tailed)	
	Wilcoxon	Smirnov
distance travelled per person per day	0.347	0.567
kilometre per journey	0.541	0.615
number of journeys per person per day	0.042	0.102

Although the p-value of the test of Wilcoxon for the number of journeys per person per day is almost significant, there is no evidence against the null hypothesis. The medians are equivalent and the means are almost equivalent for this parameter (Table 6). Finally, the p-value of the test of Smirnov for this parameter is quite large.

6. Conclusions

In order to test the effects of an informed consent letter on response rates and times and mobility parameters of the NTS, an experiment was conducted. Logistic regression analysis and event history analysis were applied to test the hypothesis that the informed consent letter has no effect on response rates respectively response times. Two parameter-free tests (Wilcoxon and Smirnov) were applied to test the hypothesis that the informed consent letter has no effect on estimates of population parameters. No significant differences between the response rates and times and the mobility parameters of the experimental group and the control group could be found. Therefore it can be concluded that there are no methodological objections to include the informed consent paragraph as a standard in advance letters of the NTS.

References

Agresti, A. (1990). *Categorical Data Analysis*. New York: John Wiley.

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. London: Chapman & Hall.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. New York: John Wiley.

Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. New York: McGraw-Hill.

Yamaguchi, K. (1991). *Event History Analysis*. California: Sage Publications.

THE MATCHING PROJECT IN SLOVENIA: WHO ARE THE NONRESPONDENTS?

Vasja Vehovar and Metka Zaletel
Statistic Office of Republic of Slovenia, Ljubljana, Slovenia

1. Introduction

The estimates based on survey samples can be biased due to various reasons. In this paper we are dealing with the nonresponse bias in official household surveys in Slovenia (Labour Force Survey, Family Budget Survey). The starting question is whether the experience from other countries (Groves 1993; Foster, Bushnell 1994) can be generalized also to the case of Slovenia. We are, of course, also interested in demographic characteristics of nonrespondents and in the consistency of results across different surveys.

The above questions were answered with the help of matched data. For this purpose, Census and registers were combined with the survey data.

The paper starts with the description of the surveys and the methodology of matching (Section 2), and continues with bivariate and then multivariate data analysis (Section 3). In some subgroups we found interesting interaction of two variables and the response rate. A logit model was also constructed for the impact of the demographic characteristics on the cooperation in the survey. We further examined the impact of individual variables on the decision to cooperate in the survey. At the end (Section 4), conclusions are made.

2. Description of the Surveys

The Data and the Procedures

We analysed the following surveys:

- *Labour Force Survey 1994* (LFS94), carried out in May 1994
- *Family Budget Survey 1993* (FBS93), annual, carried out in December 1993.

The General Social Survey was also analysed but is not presented here due to some unforeseen technical obstacles.

Fieldwork strategies were similar for both surveys (e.g. five follow-ups). Sampling plans for both surveys were also similar: they are stratified two (FBS) or three (LFS) stage surveys of households.

It is essential that households were selected through persons from the Central Register of Population of the Republic of Slovenia (CRP). Of course,

larger households had larger probability of being selected and afterwards this effect was neutralised by weighting.

In surveys analysed, either substitute units (FBS 1993) or weighting of respondents within the groups (LFS 1994) were used. We do not discuss these issues here. Some results can be found in Vehovar (1993).

Let us first define the unit nonresponse rates:

- nonresponse rate = number of nonrespondent/number of eligible units
- refusal rate = number of refusals/number of contacted units
- noncontact rate = number of noncontacted units/number of eligible units
- completion rate = number of responses/initial sample size
- eligibility rate = number of eligible units/initial sample size.

There were 3482 households included in the design of LFS94. The nonresponse rate was 8.9 %, the refusal rate 5.9 % and the completion rate 88.1 %.

Fieldwork lasted one month with beginning in May 1994. The average length of an interview was about 20 minutes. The survey mode was face to face paper and pencil interviewing. About 130 interviewers and 13 field supervisors were involved in conducting of data collection.

There were 4566 households included in FBS 1993. We have records about reasons of nonresponse for 4293 households included in the first phase of the survey in May 1993. So we can only analyse the nonresponse rate (24.6 %) and the refusal rate (11.9 %) based on first wave nonrespondents. The completion rate was 71.6 %.

Fieldwork lasted three weeks with beginning in December 1993. The average length of an interview was about 80 minutes. The survey mode was the same as in LFS 1994. There were 109 interviewers and 30 field supervisors involved.

Table 1. Basic information about analysed surveys.

	LFS 1994	FBS 1993
initial sample size	3482	4566
nonresponse rate	8.9 %	24.6 %
refusal rate	5.9 %	11.9 %
noncontact rate	2.1 %	4.0 %
eligibility rate	96.6 %	95.0 %
completion rate	88.1 %	71.6 %

The Matching

We used the following three sources:

- 1991 Census of Population, Dwellings, Households and Farm Economies, conducted by the Statistical Office of the Republic of Slovenia in April 1991
- income tax records file for 1993
- Register of Unemployed Persons in the Republic of Slovenia (May 1994).

The key variable that allowed us to combine the samples with the above data bases was the personal identification number (PIN). At the beginning we had all PINs only for persons included in LFS. In order to combine the data from FBS we first had to combine the survey data with the Central Register of Population (CRP), so that we attached their PINs. Due to many technical problems[1] we successfully combined 77 % of persons in FBS 1993 and the number of eligible persons for matching decreased to 3484 for FBS 1993.

In Table 2 we can observe the loss of data because of unmatched cases.

Table 2. Loss of data during the matching process.

	LFS 1994	FBS 1993
initial sample size	3482	4566
number of eligible households	3364	4338
number of respondents	2926	3270
number of eligible persons with PINs	3120	3484
number of persons with Census data	3047	3396
number of persons with income data	2025	2339

We lost about 2.5 % of each sample during the matching with the Census data base. Most of them (161 all together) are people who moved from abroad.

A specific problem appeared in combining the selected persons with the income tax records data base. There, a large number of persons have no income of their own and are therefore without any income tax records. For persons who were not in the income tax data base we do not know whether they do have income – but the match was not performed properly – or they do not have it at all. In any case we classify them into the lowest income category. In analysing the influence of the height of income on response or refusal, we analyse both the data with and without those cases. The analysis was run also on cases without income data.

Variables Analysed

Among variables included in the analysis, are:

From Census data:

- *age of the person* (younger than 70 years; 70 years and over)
- *sex of the person*
- *education* (less than 12 year of education; 12 years or more of education)
- *size of the household* in which the person lives (single household; larger household)

[1]The problem appeared because only PINs were recorded correctly but not name, surname and address, which were the key data in combining with CRP and might have changed for certain persons.

- *family type* (family with children; family without children)
- *level of urbanization* of settlement in which the person lives (rural areas; urban areas)
- *type of housing* (individual housing; multiunit building).

From income tax records:

- *gross income* in 1993 (1st and 2nd class on the tax scale; 3rd – 5th class on the tax scale).

From the Register of Unemployed Persons:

- *registered unemployment* in May 1994.

We checked – when available – the census and register data with information obtained from surveys. Only after the extensive analysis we decided on the categorization of variables described in brackets above. These are only final classifications of the values of the variables which are the most appropriate for the interpretation. The analysis with more complex scales showed there is no loss in the explanatory power.

It is worth mentioning that the last two variables are extremely important: unemployment is the target variable in LFS and income in FBS. We should also notice that some of the above variables refer to the selected person and other to the household of the selected person.

3. Analysis

We have analysed the following questions:

- whether the response rate is a function of explanatory variables
- whether there are some interactions of two variables influencing the response rate
- whether the refusal rate is also a function of the observed variables.

Nonresponse Rates Across Categories of Analysed Variables

Let us observe the nonresponse rates by categories. The percentage of refusals in total nonresponse was calculated as a ratio of the number of refusals and the number of all nonrespondents.

We split variables into two sets: variables describing households (Table 3) and variables describing selected persons (Table 4).

Table 3. Nonresponse rates in categories of variables describing households.

	LFS 1994			FBS 1993		
	non-response rate	refusal rate	% of refusals in nonresponse	non-response rate	refusal rate	% of refusals in nonresponse
Total	12.0	8.1	68.4	23.7	11.9	59.9
Household size						
single	23.3	9.33	34.15	34.6	17.02	48.19
larger	11.1	8.35	71.53	22.3	16.12	72.40
Housing						
individual housing unit	9.9	7.61	73.26	19.9	14.65	73.82
multiunit building	15.9	10.16	58.45	29.2	19.02	64.65
Level of urbanization						
rural	8.2	5.91	68.81	16.3	11.87	71.43
urban	15.6	10.76	65.91	28.6	19.52	68.97
Type of family						
family without children	15.2	8.13	72.60	26.8	16.13	74.85
family with children	10.6	9.07	57.02	21.7	16.28	60.30

Table 4. Nonresponse rates in categories of variables describing selected persons.

	LFS 1994			FBS 1993		
	non-response rate	refusal rate	% of non-response	non-response rate	refusal rate	% of non-response
Age						
under 70 years	11.5	8.05	67.74	23.4	16.74	71.78
70 years and over	17.4	11.44	62.00	25.7	14.11	54.90
Education						
less than high school	11.9	8.37	67.43	21.5	14.16	68.95
high school, college, university	12.3	8.48	65.77	30.7	20.64	73.43
Registered unemployment in May 1994						
unemployed	11.6	6.61	55.56	23.8	15.28	61.11
other	12.1	8.57	67.88	23.7	16.51	70.23
Income						
not known	12.1	7.77	60.90	27.6	16.47	59.47
1st and 2nd class	11.8	8.85	72.81	20.3	15.70	77.35
3rd -5th class	14.4	10.40	69.23	33.9	22.81	67.53

First of all, one should distinguish between nonresponse rates in Table 3 and those in Table 4. For example, the nonresponse rate 11.5 % among persons

under 70 years in Table 3 means that 11.5 % households with a person under 70 years did not respond.

Comparing nonresponse and refusal rates in both surveys, we can notice that – as expected – both rates are much higher in FBS93 than in LFS94. But comparing the percentage of refusals in the whole nonresponse count, we notice the same pattern in LFS94 and FBS93. There are a few exceptions to this rule: the percentage of refusals among single person households is 34 % in LFS94 and 48 % in FBS93.

We have already mentioned that income and unemployment are the target variables in each survey. We can observe that unemployment has no impact on the nonresponse rate. It is worth mentioning that unemployed persons refuse to cooperate less often than other persons, but are harder to contact.

On the contrary, we can observe a strong impact of income on nonresponse and refusal rates in FBS93. So income in LFS 1994 has almost no influence on refusal or noncontacts of households, but it is very important in FBS93. Asking about income thus increases the nonresponse on income – it looks we have a non-ignorable missing mechanism.

The largest differences in the nonresponse rate can be noticed in the level of urbanisation, type of housing and household size. Differences are statistically significant in both surveys. The differences in other variables, e.g. family type, education and age are statistically significant in only one survey or not at all, so that we cannot generalize them to other surveys. Below, we describe the differences between nonresponse rates and refusal rates for each of the independent variables:

The refusal rate in single households is almost equal to the refusal rate in larger households (9.3 % vs. 8.3 % in LFS94 and 17.0 % vs. 16.1 % in FBS93). However, nonresponse rates are much higher in single households. It is clear that single households are significantly more absent. With the increase in the number of visits of interviewers (from 5 to e.g. 8 visits) we may significantly decrease the nonresponse.

Households in multiunit buildings have a higher refusal rate than households in individual houses (10.1 % vs. 7.6 % in LFS94; 14.6 % vs. 19.0 % in FBS93) and higher nonresponse rate (15.9 % vs. 9.9 % in LFS94; 29.2 % vs. 19.9 % in FBS93). The differences are smaller than in the case of household size.

The nonresponse rate of households living in urban areas is almost twice as large as the nonresponse rate of households living in rural areas (15.6 % vs. 8.2 % in LFS94; 28.6 % vs. 16.3 % in FBS93). We can observe approximately the same relation in refusal rates.

In the case persons of over 70 years of age, we can not generalize results from one survey to the others. The result of FBS93 tells us that households

with elderly are more absent than other, but the result of LFS94 can not justify it.

Observing education, there is no difference between nonresponse rates and between refusal rates in LFS94. The difference in FBS93 is larger (30.7 % vs. 21.5 %). Again, we can not generalize the results.

To summarise the above findings, we can state that with the increase in the number of visits of interviewers we may significantly decrease the level of nonresponse at least in certain categories, e.g. in single households and households living in multiunit buildings.

Interactions

It is shown in the further analysis that there is an interaction between the level of urbanization, education and response in the LFS 1994 data base. It is also the only interaction that is statistically significant in the LFS data.

Table 5. The nonresponse rates according to urbanization and education in LFS94.

	EDUCATION	
URBANIZATION	less than 12 years	12 years and over
Rural areas	9.5 %	4.5 %
Urban areas	15 %	16.0 %

In urban areas there is no difference between nonresponse rates with respect to education but in rural areas the education makes a difference: the educated persons responded more often than the less educated ones. It seems that this interaction is specific for Slovenia since has not been reported in other countries.

The only statistically significant interaction in FBS 1993 data was the interaction between single households and the level of urbanization. In the bivariate table of associations (Table 3) we observed that single households have lower response rates. However, in Table 6 we see a distinctly higher nonresponse rate among single households in urban settlements. In rural settlements the size of households makes no difference to the nonresponse rate. This interaction is to be expected in other countries too.

Table 6. The nonresponse rate according to urbanization and household size in FBS93.

	SIZE OF HOUSEHOLD	
LEVEL OF URBANIZATION	single	larger
Rural areas	19.0 %	16.7 %
Urban areas	41.7 %	27.3 %

Logit Models For Predicting Response

We constructed standard logit models for predicting response for both surveys:

$$g(y_i)^{(s)} = \beta_0^{(s)} + \sum_{j=1}^{8} \beta_j^{(s)} x_{ij}^{(s)},$$

Variables that are not statistically significant for individual surveys have empty cells in the table of coefficients for corresponding surveys.

$s = 1$ for LFS94, $s = 2$ for FBS93;
dependent variable y = response (0 = nonresponse, 1 = response),
$g(y)$ is logit-link function.

Independent variables are:

$x_1 = 1$ for persons in urban areas, otherwise 0;

$x_2 = 1$ for persons with 12 years or more of education (finished secondary school, non-university or university degree), otherwise 0;

$x_3 = 1$ for persons over 70 years of age, otherwise 0;

$x_4 = 1$ for persons living in single households, otherwise 0;

$x_5 = 1$ for persons living in multiunit buildings, otherwise 0;

$x_6 = 1$ for persons belonging to the 3rd, 4th or 5th category of the tax scale, otherwise 0;

$x_7 = 1$ for persons in urban areas with 12 years or more of education, otherwise 0 – this variable represents interaction between the level of urbanization and education;

$x_8 = 1$ for persons living in single households in urban areas, otherwise 0 – this variable represents interaction between the level of urbanization and single households.

Table 7. Logit models for LFS94 and FBS93: logit coefficients and their standard errors (standard errors in parenthesis).

	LFS 1994		FBS 1993	
	logit coeff.	stand.error	logit coeff.	stand.error
Intercept	2.3722	(0.1131)	1.7780	(0.0892)
Urban areas	-0.3672	(0.1662)	-0.5927	(0.1111)
Education	0.7394	(0.3048)		
Age over 70 years	-0.3699	(0.1807)	-0.4645	(0.1952)
Single households	-0.6241	(0.1998)		
Multiunit building	-0.2394	(0.1423)		
Income			-0.6098	(0.1521)
Interaction education/urban	-0.8731	(0.3383)		
Interaction single/urban			-0.7072	(0.2575)

The highest logit coefficient we can find in LFS 1994 is the coefficient for interaction education/urban. The coefficient of –0.8731 means that persons with higher education living in urban areas have probability of response 82 %. Persons, living in rural areas in individual houses with higher education, of age under 70 years and not living in single households, have the highest probability of response: i.e. 96 %.

We can not generalize the result of the logit analysis of LFS94 on FBS93: people, living in urban areas, younger than 70 years, with lower income and not living in single person households, have the highest probabilities of response (85 %). We can find the lowest probability of response (38 %) for people living in urban areas, in single households, older than 70 years and with higher income.

Relative Bias of the Estimates

For variables that we had at our disposal we were able to calculate the relative nonresponse bias:

$$B = \frac{n_{nr}}{n} \cdot \frac{(y_r - y_{nr})}{y} \cdot 100,$$

where:
n_{nr} is the number of nonrespondents in the sample, n is the total sample size, y_r is the value of the estimator y for respondents, y_{nr} is the value of the estimator y for nonrespondents, and y is the value of the estimator in the total sample. The relative bias (B) can be interpreted as the percentage of error when the estimate is based only on respondents.

We can observe the results in Table 9.

Again, let us repeat the results of target variables of our surveys – unemployment rate and average income. The estimate of registered unemployment has no nonresponse bias in LFS 1994. Income estimate is biased in FBS 1993 (5 %), which means that the estimates in FBS 1993 connected to income are underestimated by 5 % only because of nonresponse. Bias in the estimate of income in LFS 1994 is less than 0.5 %, and thus negligible. Here we have an example of a non-MAR missing process (Little and Rubin 1987): asking about income (FBS) makes people with a certain (high) income more uncooperative, but not asking them (LFS) keeps them cooperative.

Table 9. Relative nonresponse bias in LFS94 and FBS93.

		Total	Nonrespondents	Respondents	R Bias
% of single households	LFS94	5.9	11.6	5.2	-13.2
	FBS93	7.5	11.2	6.4	-15.1
% of people living in multiunit building	LFS94	32.3	43.2	30.8	-4.6
	FBS93	35.2	44.3	32.5	-8.0
% of people living in urban areas	LFS94	52.2	67.0	50.2	-3.9
	FBS93	56.8	69.7	52.9	-7.0
% of families with children	LFS94	72.1	64.3	73.2	1.5
	FBS93	69.3	64.8	70.7	2.0
% of people older than 70 years	LFS94	9.6	13.8	9.0	-6.1
	FBS93	11.9	12.9	11.6	-2.6
% of people with 12 years or more of education	LFS94	33.1	33.7	33.0	-0.3
	FBS93	15.2	20.4	13.7	-10.5
average household size	LFS94	3.7	3.4	3.8	1.4
	FBS93	3.6	3.3	3.6	2.0
average apartment size	LFS94	75.2	73.1	75.5	0.4
	FBS93	73.3	70.8	74.1	1.1
average number of rooms in the apartment	LFS94	2.9	2.8	3.0	0.7
	FBS93	2.9	2.8	2.9	0.7
% of unemployed persons	LFS94	8.3	8.0	8.4	0.5
	FBS93	6.5	6.6	6.5	-0.1
average income (in 1000 Slovenian Tolars)	LFS94	619	638	617	-0.4
	FBS93	671	774	642	-4.7

Among demographic variables the estimate of the percentage of single size households is especially biased. The estimates in FBS 1993 and LFS 1994 are both underestimated by around 15 %. Also underestimated is the education in FBS 1993 and LFS 1994, where we underestimated the percentage of persons with secondary school, non-university and university degree by 10 %.

4. Conclusion

We can conclude the following :

Certain characteristics of nonrespondents are the same as in other countries, e.g. nonrespondents are more likely to be older, living in single households and in urban areas (Groves 1993).

However, some of the characteristics are country specific, e.g. interaction of education, level of urbanization and response.

Nonresponse bias was found to be very high (Rbias = 5 %) for the income variable, but surprisingly low for the unemployment rate.

We should remind of some methodological problems with the interpretation:

Some variables describe selected persons (e.g. income) and other their households (e.g. family type).

There is a certain time lag between census and our surveys and some information about selected persons are out of date. Fortunately, there is no time lag in both key variables we are most interested in – income and unemployment.

Also, we must be aware of a relatively large loss of data due to no PINs in the case of FBS 1993 which can, of course, create certain problems with the interpretation.

However, we believe that the main conclusion are robust with respect to the above described limitations.

With demographic characteristics of respondents and nonrespondents we can predict relatively well the response in household surveys (70 % in logit analysis). Knowing these predictors we can now be more efficient in using some techniques for reducing nonresponse.

It is obvious that the most interesting question was not yet addressed here; i.e. whether weighting and imputation can remove the nonresponse bias. The extent of such improvements is definitely the issue of the future work on this matched data.

References

Allard, B. and Dufour, J. (1994). *Characteristics of Respondents and Nonrespondents in the Canadian Labour Force Survey*. Statistics Canada.

Brehm, J. (1993). *The Phantom Respondent: Opinion Surveys and Political Representation*. Ann Arbor, Michigan: The University of Michigan Press.

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

Fingleton, B. (1984). *Models of Category Count*. Cambridge: Cambridge University Press.

Foster, K. and Bushnell, D. (1994). *Non-response Bias on Government Surveys in Great Britain*. OPCS, Great Britain.

Gilbert, N. (1993). *Analysing Tabular Data*. London: UCL Press.

Goyder, J. (1988). *The Silent Minority*. New York: Wiley.

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

Groves, R. M. and Couper, M. P. (1993). *Correlates of Nonresponse in Personal Visit Survey*.

Kish, L. (1965). *Survey Sampling*. New York: Wiley.

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

Vehovar, V. (1993). Field Substitutions in Slovene Public Opinion Survey. *Contributions to Methodology and Statistics. Ljubljana: FDV*. 39–66.

TRENDS OF NONRESPONSE IN THE SURVEYS CONDUCTED BY STATISTICS FINLAND

Kari Djerf
Markku Heiskanen
Pertti Kangassalo
Seppo Laaksonen
Paavo Väisänen
Statistics Finland, Finland

PREFACE

This paper is a first attempt to assess the general trends in the survey climate in Finland. Naturally we have kept a close eye on the latest developments in nonresponse, but so far we have not combined data from various surveys into one publication. Even now the contents cannot be regarded as comprehensive: we included some examples of surveys with long time series (e.g. the Labour Force Survey), surveys with diaries (the Household Budget Survey, the Time Use Survey), some more infrequent surveys and examples of surveys financed by customers. One survey on farming was also included. But we still lack many others, especially enterprise surveys. These should be analysed in a similar manner in future.

We have mostly used descriptive analysis on the nonresponse figures here. However, our data should be combined with information on attitudinal changes among the population. As a survey organisation we certainly have posed at least some questions which could be useful, but were not analysed in the present context. Therefore, we must restrict ourselves here to the historical trends.

One can recognise that nonresponse has grown since the 1980's. In the Labour Force Survey (LFS) the annual nonresponse rate rose by about 2 percentage points until the 1990's. Similar changes occurred also in the Income Distribution Survey and the Household Budget Survey (HBS). Therefore, the survey climate must also have changed somehow, even though these changes had not been that drastic.

However, one should not neglect other changes either. They might be due to our own or someone else's actions. For example, at Statistics Finland

we have merged surveys. As regards the Labour Force Survey, it is common that some surveys financed by customers are being added to the LFS. Normally these additional surveys are not too burdensome. Therefore, the nonresponse rates were not increased but slightly. Recently we have had to change our data collection strategy to comply with the Eurostat requirements: we use a sample of individuals, whereas the sample of the harmonised LFS consists of households. In the spring of 1995 that simple action nearly doubled the refusal rate.

On the other hand, we have combined the Income Distribution Survey with other surveys practically every year since 1988. The combination is reasonable in one important respect: the contents of the survey make an integrated whole. Likewise we have changed the data collection design of the Household Budget Survey from 1994 on. In these two complicated surveys the non-response rates increased by more than 5 percentage points most probably due to our own measures.

An even more severe change took place in the Farm Survey when the tax authorities refused to include certain important information in the tax forms in connection with the reform that was introduced. The sudden attempt to include those questions in our Farm Survey was a disaster. As regards that particular additional questionnaire nonresponse rose to nearly 50 per cent. These examples may not be the only ones.

We can rather easily try to improve our own ways of thinking and making rapid changes (those which should in no case have any kind of effect on the results). It needs only more careful work. But it is really demanding to try to influence some other officials.

The survey organisation and data collection strategy are the most important tools for seeking good results. After data collection we can still try to attain as good results as possible. At Statistics Finland we have rather often used reweighting techniques in estimation. Those methodological studies are, however, not the main subject of this publication. Therefore, a reader interested in them should look at the list of relevant literature published by Statistics Finland.

This report is a joint venture. Markku Heiskanen have described various surveys (Survey on Living Conditions, Drinking Habits, Sexual Behaviour and Victimisation). Pertti Kangassalo took charge of the Household Budget Survey. Seppo Laaksonen's contribution is on the Time Use Survey. Paavo Väisänen wrote about the Farm Survey and I am responsible for the rest.

Kari Djerf

CONTENTS

LABOUR FORCE SURVEY

Design

The Finnish Labour Force Survey (LFS) is essentially a survey on individual persons, not on households. The target population consists of the so-called economically active population which has been defined as the resident population aged 15–74 years. The interviews are rather short; the first one only about 15 minutes and the subsequent interviews even shorter. Therefore, it is little wonder that it is used as a host to many other surveys, primarily to those financed by customers.

Data are collected mainly by telephone interviews (94 per cent). So far we have used PAPI questionnaires but now we are starting with decentralised CATI. Some 15 hundred interviews are annually conducted by the centralised CATI facility. The active work time is short, only two weeks. We measure the labour market status in the very week where the 15th day of the month happens to be. Interviews can be started the following week.

We use a rotating panel design where each person is interviewed five times over a period of 15 months. The lag between the interviews is three months, except once when it is six months. However, elderly people (aged 64–74 years) are interviewed only every second time. Meanwhile, information on their activities is imputed from the earlier interviews. Also proxies (i.e. eligible family members) are allowed but not in the first interview. The share of proxy interviews is about 4 per cent.

The first wave sample is drawn from the Central Population Register by systematic sampling. We regard it as an approximation of the simple random sampling because we have not found dependence between the order of the frame (dwelling code) and the study variables. The order only guarantees that the sample is geographically representative. So far we have used fairly detailed post-stratification (sex, age, region), but discussion has also been raised whether to use other auxiliary information.

The Harmonised Labour Force Survey of the European Union has been merged to our old design since 1995. There is a major change in design since the harmonised survey needs information on the entire household. Therefore, the design has been adopted to meet the standard by Eurostat in the three months in the spring.

Nonresponse

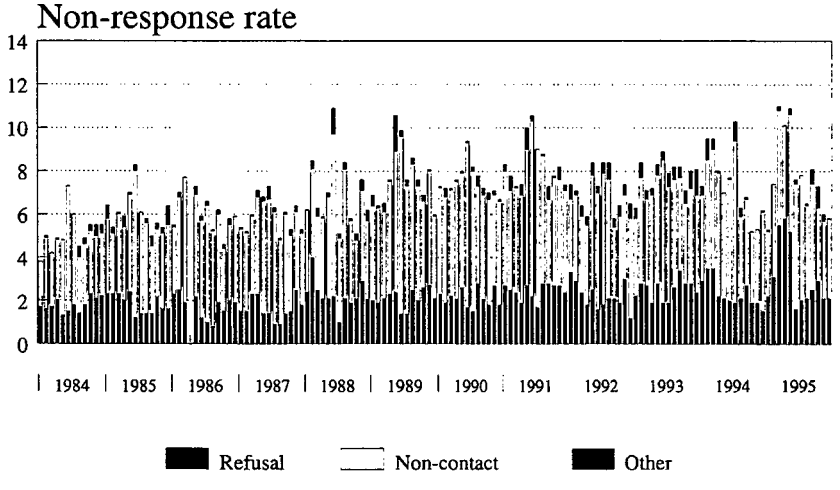
Here we refer to the figures of the first wave because the data are clean from the following disturbing features:

a) proxies and imputations are allowed in the subsequent interviews.
b) panel attrition may be a problem (However, the sampling procedure takes this, to some extent, into account, because people are actually sampled after reaching the age of 14 and they will be included once they are 15. Those reaching the age of 75 are respectively removed from the sample).
c) non-contact cases are reissued in the latter waves.

In the LFS the nonresponse rate has grown during the years. Since 1984 the annual average has varied between 5.2 and 8.3 per cent. In the 1980's it rose gradually from 5 per cent to about 7.5 per cent where it has remained ever since. The patterns of nonresponse have remained fairly stable: non-contact problems dominate in the first wave. For example, in 1995 the non-contact rate was 4.5, the refusal rate 4.1 and other reasons accounted for 0.2 per cent out of a total of 7.8.

One can easily see that the nonresponse rate grows systematically in the summer months, especially in July. We have contact problems during those months because a lot of families have a summer cottage where they stay for months.

Fig. 1. The monthly nonresponse rates of the LFS from 1984 to 1995.



An alarming feature is detected in the spring months of 1995. The Harmonised European Labour Force Survey raised the nonresponse rate by about 3 percentage points. The importance is even more clearly seen in the figure describing the refusal rate. Thus changing over to interview the entire

household caused us a real loss in information because we will obtain the higher nonresponse rates also in the subsequent interviews.

Fig. 2. The monthly refusal rates of the LFS from 1984 to 1995.

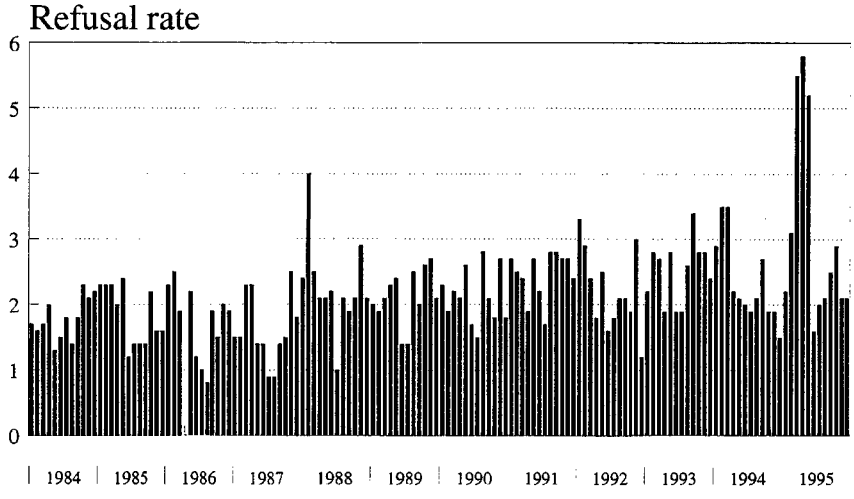
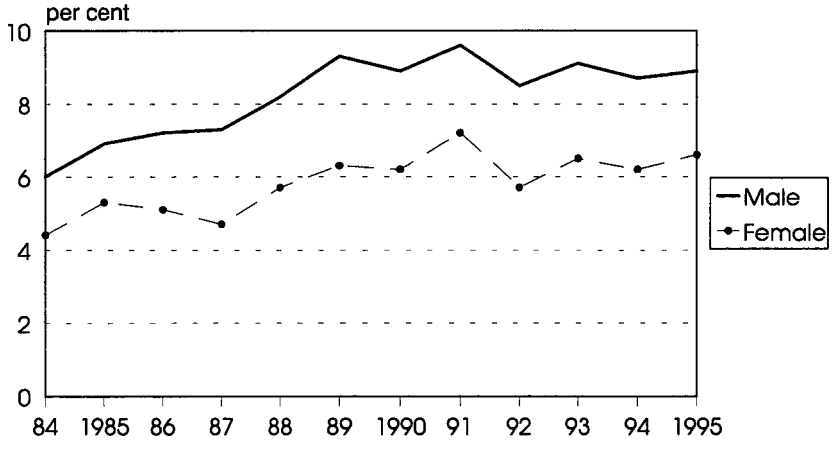
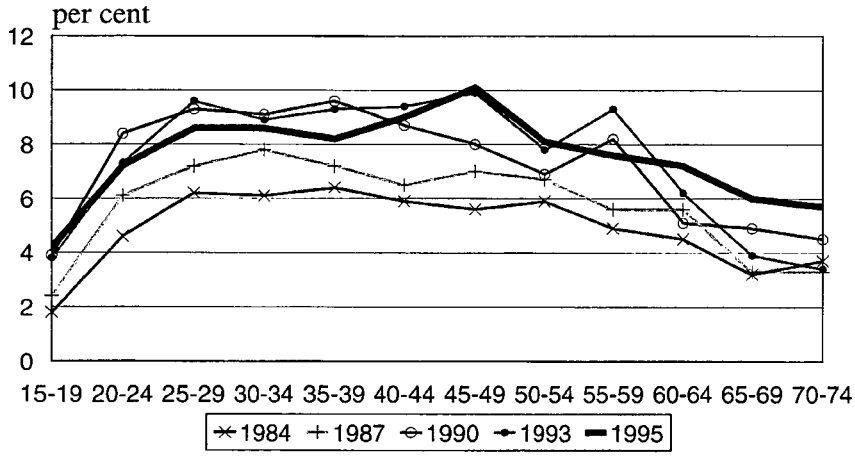


Fig. 3. Nonresponse rates of the LFS by sex, 1984–1995.



There is a stable relation between the sexes: women always have a couple of percentage points higher response rates than men (e.g. 93.4 versus 91.1 in 1995). As regards the age, one can see a flat inverse U-surface in nonresponse, the most reluctant interviewees being in their active working age.

Fig. 4. Nonresponse rates of the LFS by age in selected years.



In the geographical consideration two regions tend to have a higher nonresponse rate than the rest of the country. These are Greater Helsinki and the Autonomous Territory of the Åland Islands. In 1995 the nonresponse rates of these regions were 12.6 and 16.5 per cent, respectively, whereas in the Eastern and Central Finland the rate was generally less than 6 per cent.

An even more interesting feature was found when the survey data were merged with the register information. From the Register of Unemployed Job Seekers maintained by the Ministry of Labour you can, for example, learn whether the person is officially a job applicant or not. The rate of unemployed job seekers happened to be double among the nonrespondents as compared to the respondents. This is an indication of bias and therefore we must consider taking this information into account in estimation.

Panel attrition

The panel structure is rather short which helps us to maintain all the subsamples in good condition. A fair method of calculating the nonresponse rate over the whole span of the data collection period must leave all imputations and other technical tricks out. Such a calculation reveals, however, that panel attrition does not seem to be a serious problem.

Table 1. Nonresponse rates of the LFS over the entire study period. Rates are calculated on the basis of the real contact attempts (no imputations neither proxy interviews).

Start	Wave One (+3 mo.)	Two (+6 mo.)	Three (+12 mo.)	Four (+15 mo.)	Five
Jan. 1994	7.3	8.1	7.8	7.5	7.5
Feb. 1994	9.5	7.7	7.7	7.6	8.1
March 1994	9.5	9.3	7.2	8.0	8.7
April 1994	8.0	8.1	6.0	7.0	6.8
May 1994	7.0	8.1	6.7	8.9	8.0

HOUSEHOLD BUDGET SURVEY

The Household Budget Survey is the largest of the surveys conducted by Statistics Finland. It has been carried out since 1966. The nonresponse rate in the HBS has varied a lot throughout the years. The lowest rate (after diary-keeping) has been 22 % and the highest 38 %. Because of the poor survey atmosphere mainly due to the economic recession, the nonresponse rate has risen up to 35–37 % in the 1990's. Furthermore, nonresponse in the HBS has traditionally been very skewly distributed according to certain important variables, e.g. size of household, area, education, and socio-economic group. Certain procedures for reducing nonresponse have been applied. In addition, the consequences of nonresponse have been adjusted by using an estimation method.

1. Some Aspects of the Survey

Purpose and Frequency The main purpose of the Finnish Household Budget Survey (the HBS) has traditionally been to provide information on the structure of the consumption expenditure of households for use in the revision of weights when compiling the consumer price index. In addition to consumption, the HBS has been used to examine the income, purchasing patterns and ownership of consumer durables of households as well as social benefits in kind.

From 1966 to 1990, the HBS was carried out about every five years. In 1994 Statistics Finland introduced an annual HBS with a reduced sample size. (The data for the three following years – from 1994 to 1996 – will be combined later to get a larger data.) The sample was selected separately for each survey. The content and survey method also changed somewhat from

one survey to another. Statistics Finland has also prepared standardised time series data files from the surveys for 1966–1990.

Sampling The population of the HBS consists of all Finnish households and their members. Institutionalised persons are excluded. For the 1981–1990 surveys about 12,000 persons from different households in the target population were selected from the Central Population Register by geographically stratified sampling. For the early years (1966–1976) there were two partly different samples, one for diary-keeping and one for the yearly interview. For the annual survey (from 1994 on) about 3,500 persons are selected each year by geographically stratified sampling.

Data Collection The basic data for household surveys have been compiled in the HBSs partly by employing field research methods and partly directly from administrative registers by means of ADP. In the surveys from 1981 up to 1990 diaries were kept fortnightly by more than 8,000 households, divided over all two-week periods of the year. In these surveys two interviews (PAPI) were arranged, an initial interview before the recording period and a yearly interview (end-of-the-year interview) with the whole of the preceding year as a reference period.

In the 1966–1976 surveys the first one of the two samples took part in diary-keeping (incl. initial interview) and the second one in the yearly interview. In those surveys the diary-keeping period was four-week long.

In the annual HBS the data collection method is very similar to the one used in the 1981–1990 surveys. There are, however, two differences: only one interview is arranged and the method used is CAPI. The interview is conducted before the recording period. To lighten the load of the interview income items are no longer asked in the interview, but will be taken from registers later. The quantities of foodstuffs are no longer to be recorded either.

2. Nonresponse

Nonresponse in the Finnish HBS has traditionally been very skewly distributed according to certain important variables. In the following, nonresponse and its structure throughout the years are dealt with.

Magnitude of Nonresponse

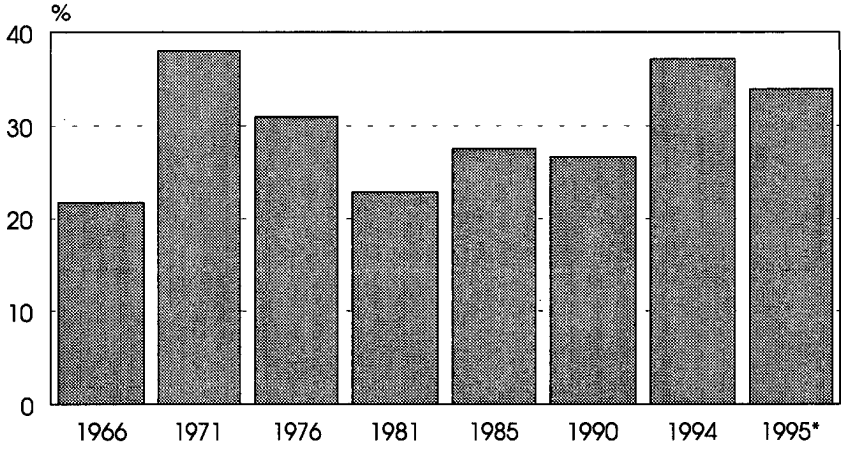
Table 1 shows the nonresponse rates in the 1990 HBS and how the final data were composed.

Table 1. Composing of data in the 1990 HBS.

	N	%
Gross sample	12,053	
– Over coverage	297	
Net sample	11,756	100.0
– Nonresponse in the initial interview	2,254	19.2
Accepted initial interviews	9,502	80.8
– Nonresponse in diary-keeping	866	7.4
Accepted data after diary-keeping	8,636	73.5
– Nonresponse in the yearly interview	378	3.2
Accepted final data	8,258	70.2

Due to changes in the way how the HBS is conducted it is difficult to make nonresponse comparisons between all the survey years. However, the nonresponse rates after the initial interview and diary-keeping are displayed below.

Fig. 1. Nonresponse rate after diary-keeping in the 1966–1995 HBSs.*



In the old surveys for 1966, 1971 and 1976, 3,300–3,900 households in diary-keeping and 4,800–8,000 households in the yearly interview provided fully acceptable data. The nonresponse rate in diary-keeping varied from 22 up to 38 %, and in the yearly interview between 9 and 23 %. In the 1971 HBS nonresponse was clearly at its highest. The reasons for the poor response rates in the HBSs of the 1970's were deemed to be the heaviness of diary-keeping, the problems with making contact with households especially in big cities and critical public opinion. In addition to this, the interviewer organisation was new and undeveloped in those days.

In 1981–1990 the data consisted of 7,400–8,300 households and the overall nonresponse rate was 25–30 %. Of that, 23–27 percentage points was

nonresponse in the initial interview and diary-keeping and the rest was nonresponse in the yearly interview.

Reducing nonresponse was set as a special aim in the HBS of 1981. For that purpose, the burden of diary-keeping was lightened (a two-week period), contacting the households was made easier, interviewers were trained and the survey was announced through media. In 1990 better results were obtained due mainly to the favourable atmosphere towards the survey and interviewers' hard efforts. Without the substitution test for one person households in some areas the response rate would, however, have been somewhat poorer.

In the two latest surveys for 1994 and 1995 the data consisted of 2,200–2,300 households. The overall nonresponse rate rose up to 37 % in 1994. In the ongoing 1995 survey nonresponse rate seems to settle down at 34–35 %. The economic recession of the 1990's has obviously had a negative impact on the survey atmosphere.

Reasons for Nonresponse

Refusing to respond to the interview has clearly been the most significant reason for nonresponse in the HBS. In general, refusals make up over 80 % of (interview) nonresponse (see Table 2). In the old surveys also removals and antiquated sampling information were important reasons for nonresponse.

Table 2. Distribution of nonresponse by reason in the 1994 HBS.

	N	Nonresp. %
Refusals	828	23.9
Was not reached	161	4.6
Diary nonresponse	280	8.1
Other reasons	15	0.4
Total	1,284	37.1

Structure of Nonresponse

Size of Household Proportionately, the majority of non-respondents are found among the smallest households in the Finnish HBS. Nonresponse among one person households has normally amounted to about 40 %. Willingness to participate increases with the size of the household. On the other hand, the response rate among the biggest households may be somewhat poorer again, which can be explained by the heavy burden of keeping diaries. (See Figure 2.)

Fig. 2. Nonresponse rate by size of household in 1981–1995 HBSs.*

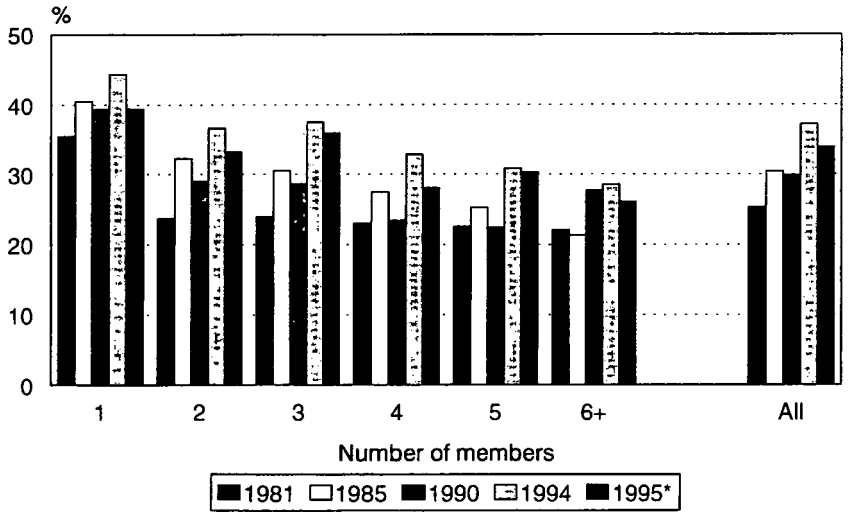
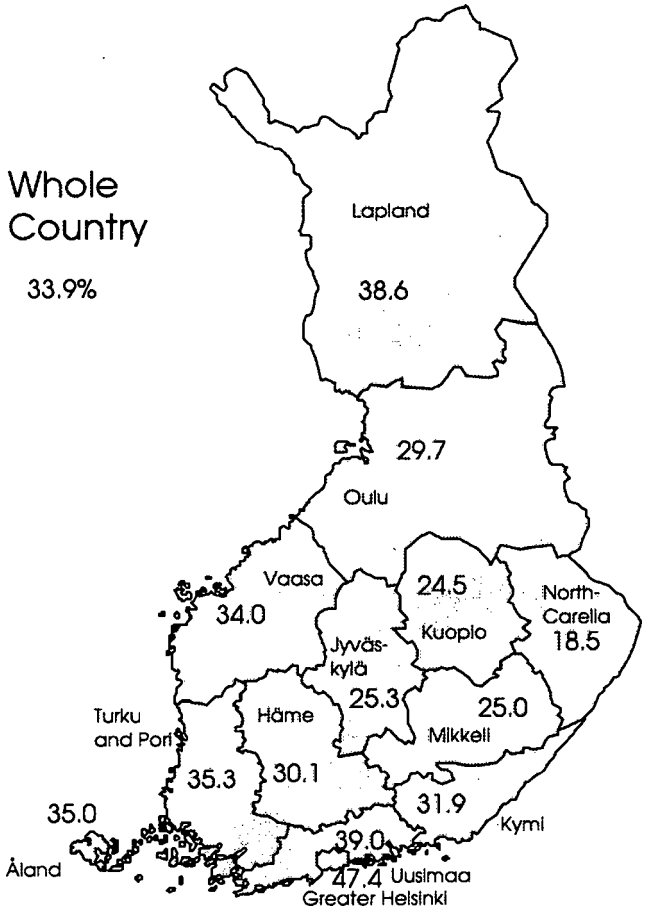


Fig. 3. Nonresponse rate by province in the 1995 HBS.*



Areas The distribution of nonresponse according to area has been very skew in the HBS. Figure 3 shows that the best results have been achieved in the eastern and central parts of Finland. The Swedish speaking Åland is in its own class as regards nonresponse: generally every other household fails to respond there (but exceptionally not in 1995). In the southernmost part of Finland, particularly in Greater Helsinki, nonresponse amounts to nearly 50 %.

In cities the willingness to respond is at a much lower level than in rural areas. The average nonresponse rate has been as high as 40 % in cities, whereas in the country it has been only 25. In the cities of southern Finland even more than 40 % of households fail to respond. Correspondingly the smallest rates in the rural areas of central and northern Finland remain below 20 %. All in all, the more densely populated the area, the more probable the inclination to nonresponse. A typical non-responding household lives in a city in southern Finland and is a one-person (retired) household. (See Table 3.)

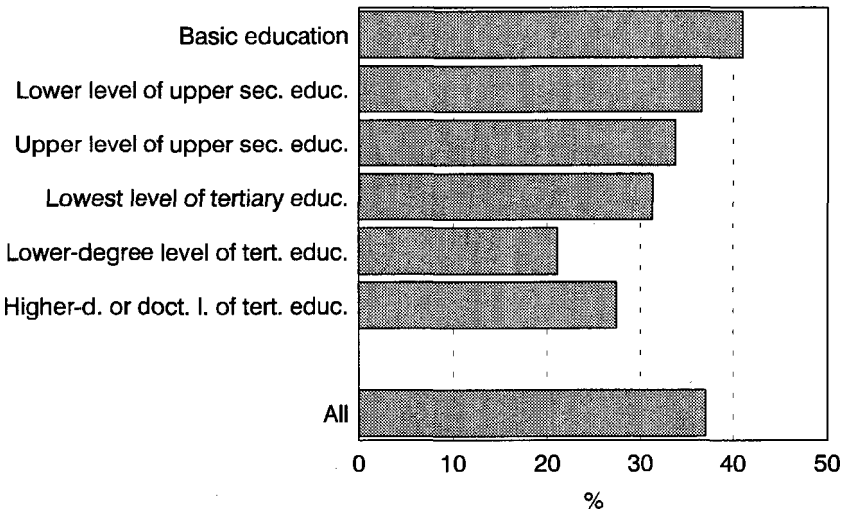
In the 1990 HBS we succeeded in lowering the nonresponse rate below 40 in Greater Helsinki with the help of a substitution test. Substitution was also used in some other big cities. Other procedures for reducing nonresponse were also applied.

Table 3. Distribution of response and nonresponse in the 1985 HBS.

Major area	Response	Nonresponse	Total
Greater Helsinki	18.8	26.9	18.9
The rest of southern Finland	43.7	46.0	44.3
Cities	26.9	31.5	28.1
Rural municipalities	16.8	14.5	16.2
Central Finland	26.0	17.2	25.4
Cities	12.2	9.9	12.3
Rural municipalities	13.8	7.3	13.1
Northern Finland	11.5	9.9	11.4
Cities	5.3	5.3	5.3
Rural municipalities	6.2	4.6	6.1
Whole country	100.0	100.0	100.0

Income and Education There are not so big differences in willingness to respond by income level as there are by area and size of household. Proportionally, the number of non-respondents is generally the greatest at both ends of income distribution, at the very lowest and the very highest level. But, evidently, the higher the education level, the lower the nonresponse rate in the HBS (see Figure 4).

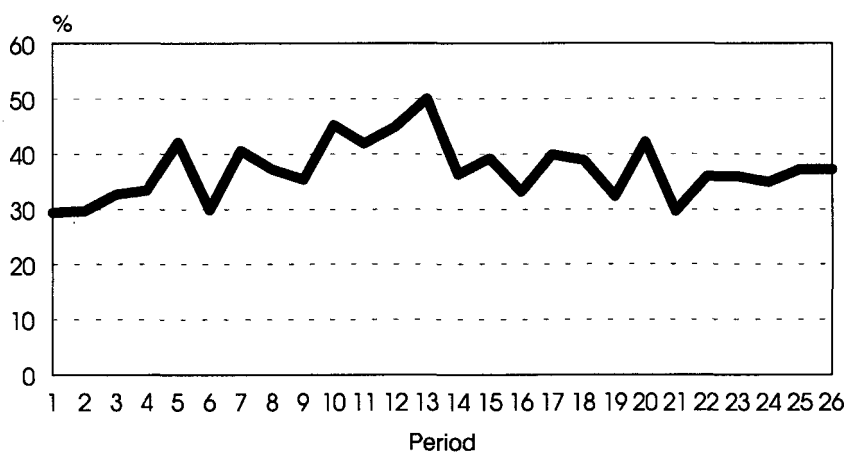
Fig. 4. Nonresponse rate by level of education in the 1994 HBS.



Socio-economic Groups As far as the socio-economic group is concerned, nonresponse among persons without a profession has usually been as high as 50 %. This group contains mostly retired people and students. The lowest nonresponse rates are found among the self-employed in agriculture, often under 20 %. Of the other self-employed, the share of those not responding is usually twice as high.

Diary-Keeping Periods Figure 5 shows that the nonresponse rate varies a lot according to diary-keeping period. Summer-time, Christmas and Easter holidays in particular, are problematic in this sense. Towards the end of the year interviewers' tiredness usually causes the nonresponse rates to rise.

Fig. 5. Nonresponse rate by two-week diary-keeping period in the 1994 HBS.



3. Adjustments for Nonresponse

Since 1981 structural differences between the sample and the population caused by nonresponse or other factors have been corrected in the HBSs by an estimation method (Ekholm and Laaksonen 1991, Laaksonen 1988 and 1992, Djerf and Lindqvist 1993). Thus the weighted sample better corresponds to the distribution of the population geographically, by the size of household and property income. It is, however, impossible to attain perfect alignment with respect to the population in this way.

INCOME DISTRIBUTION SURVEY

Data for the Income Distribution Survey (IDS) have been collected annually since 1977. Before 1977 income data were, however, obtained from household expenditure surveys. From 1977 to 1981 data were collected by using mail surveys and thereafter by interviewing.

Data are collected from various sources. The vast majority of income variables are obtained directly from administrative registers the most important of which are the taxation register and the pension registers. Register information covers more than 90 per cent of the disposable income. Nevertheless, some income items (e.g. interest payments on deposits) are not included in any official source and they must be asked by interviewing. And what is most important, the composition of the household can only be decided on the basis of the interview information.

A note to the reader: Due to retrospective nature of data collection there is a timing problem: the true statistical year is always one year behind the actual year. For example, data for the 1994 IDS were collected in 1995. Please, add one year to the IDS when comparing the figures with other surveys.

Sampling design

Since the early 1980's we have used a two-year rotating panel design in the IDS. The Taxation Register was used as a sampling frame for several years. Frame aging was a real problem with the rather fine-tuned selection procedure (one variant of the so-called snowball sampling). At present, we use a two-stage sampling design. First we take a rather large master sample from the Central Population Register where the inclusion probabilities for households reflect the household size (number of persons aged more than 15

years in each dwelling unit). The master sample is then merged with the most recent taxation records and a preliminary socio-economic class is derived according to the person with the greatest taxable income. The sample is finally drawn using stratification according to socio-economic class; farmers, entrepreneurs and high-income wage earner households having higher inclusion probabilities.

The basic weights are naturally obtained by using the reciprocals of the inclusion probabilities. Previously post-stratification was employed – most probably for fighting nonresponse bias. Unfortunately the regional post-strata did not yield enough good results. Nonresponse adjustment methods were introduced for the 1986 data (see Laaksonen 1992a). Both model-based and weighting class adjustments have been used thereafter.

In 1994 we combined the nonresponse adjustment with calibration. Calibration was performed separately for each panel wave by using the following variables:

- aggregate taxable income subject to state tax from earnings, pensions etc.
- aggregate taxable income subject to state tax from property
- aggregate taxable wealth
- population distribution by sex and following age categories: 0–4, 5–9, 10–14, 15–19, 20–24, 20–39, 40– 54,55–64, 65–74, 75–.

Calibration improved the correspondence of many income estimates with the known totals, and the population structure is now correct. (see Djerf 1995b).

Nonresponse

Nonresponse rates have varied between 12 and 25 per cent. In the past few years it has been about 22 per cent or more. The real reason for increased refusals in 1987, 1989, 1993 and 1994 is most probably related to respondent burden. The IDS has also served as a platform for many other surveys. In 1987 data were collected for the Study on Household Saving and Indebtedness, in 1989–1990 for the Household Budget Survey, in 1993 for the Survey on Living Conditions (SLC) and in 1994 for the Survey on Wealth (SW). The idea to combine the IDS with some other survey is rather sensible. The plain IDS interview is fairly short (22 minutes), but unfortunately the contents are disjoint. A respondent will hardly understand how his or her responses will be related to income distribution. Better survey contents could be achieved by combining the IDS with another survey the topic of which is related to income or social affairs. The merging increases the nonresponse rate by about 6 to 7 percentage points. However, nonresponse in the second wave has stayed rather stable, at 5 per cent on average.

Like in many other surveys, the geographical distribution of nonresponse is fairly skew. It is highest in Greater Helsinki and in other "big cities", whereas in the eastern provinces it is less than 20 per cent. The household size is another important factor. The bigger households tend to participate much better than the smaller ones (the figures are counted for the so-called register households which are used in sampling).

Fig. 1. Nonresponse rate of the IDS, 1983 –1994.

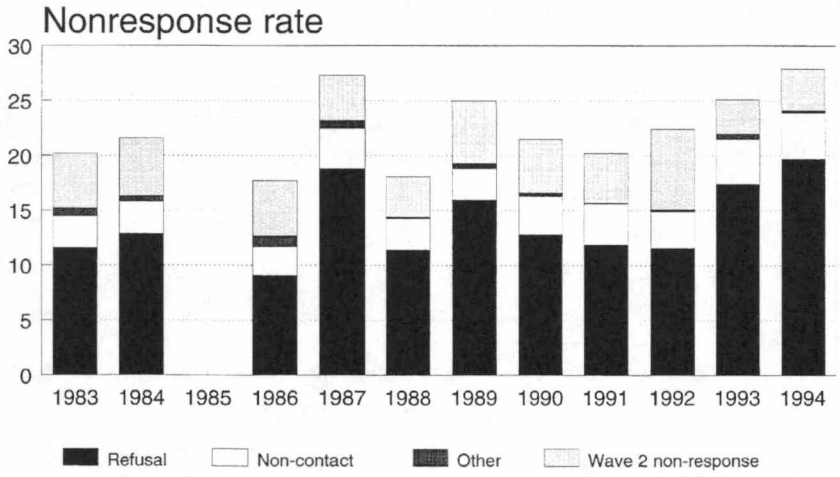
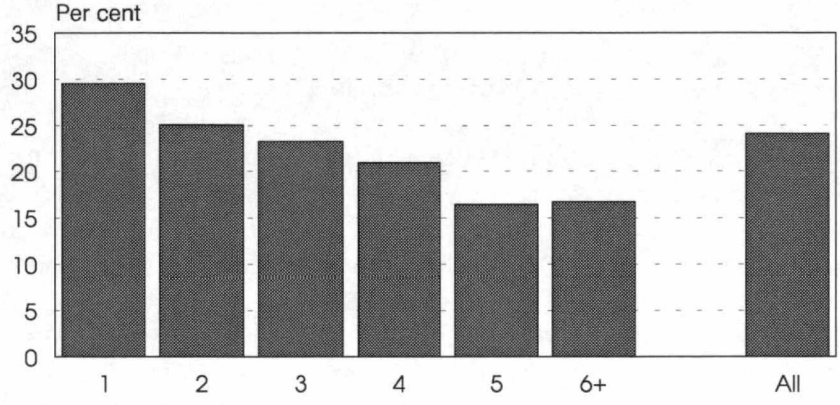


Fig. 2. Nonresponse rate according to household size, IDS 1994 wave 1.



The socio-economic stratification (used in sampling phase) also reveals some interesting patterns in nonresponse. Traditionally one could have noticed a tendency that the upper-class employees (civil servants and other office workers) on one hand and farmers on the other hand often have higher response rates than other socio-economic groups. Both very poor and very rich people refuse more often than others. That seems to be the case in the 1994 survey (data collected in the spring of 1995) when we were asking questions on wealth. A couple of interesting features can, however, be seen

from the table below. The Survey on Wealth may have some negative impact on the response rate among the rich. In general, the households in the higher income categories were more reluctant to participate when we compare the results with those of 1993 when the Survey on Living Conditions was merged with the IDS. Especially the higher income entrepreneurs refused very easily. Thus the topic of the survey seems to affect the results.

Table 1. Nonresponse according to the preliminary socio-economic category (stratification variable) in the 1993 and 1994 Income Distribution Survey, Wave 1.

Socio-economic category	1993 %	1994 %
Wage and salary earners 1	23.2	23.1
Wage and salary earners 2	26.1	24.0
Wage and salary earners 3	23.3	20.8
Wage and salary earners 1	15.8	22.3
Entrepreneurs 1	23.7	27.1
Entrepreneurs 2	23.8	42.0
Farmers 1	17.6	16.2
Farmers 2	11.9	9.7
Pensioners 1	28.7	28.2
Pensioners 2	25.5	27.4
Other 1	25.8	27.3
Other 2	18.0	24.9
Not pre-classified (children etc.)	27.1	24.0
All	23.5	24.1

THE 1994 SURVEY ON LIVING CONDITIONS

Statistics Finland has so far conducted a Survey on Living Conditions (SLC) three times: in 1978, 1986 and 1994. The results of the two latter ones are comparable at least to some degree.

Year of the study	Response rate (%) Total	Men	Refusals Total (%)	Sample size* (n) Total	Overcoverage (%)
1978	84.9	84.9	6.2	2,971	1.4
1986	86.9	86,7	9.3	13,876	1.6
1994	73.0	72.1	21.9	11,843	2.1

* overcoverage excluded

The response rate was highest in the 1986 survey. The result was probably achieved thanks to the strong efforts to fight nonresponse during the fieldwork. Local interviewer groups took care of eventual nonresponse cases.

Those teams also planned together the means to obtain interviews. The refusal rate was higher in 1986 than in 1978 probably because the initial contact rate was better: the profile of nonresponse changed from non-contacts to refusals.

In 1994 nonresponse rose sharply compared to 1986. Nonresponse rates increased in some other surveys in the 1990's, too, but mainly for different reasons. Firstly, the 1994 living conditions survey used the sample of the Income Distribution Survey of Statistics Finland, whereas the 1986 survey was a normal survey of its own. In panel surveys (such as IDS) nonresponse tends to increase at least as regards the second interview phase. Secondly, combining two surveys increased the response burden which led to refusals. Thirdly, the possibility to make the income distribution survey by telephone had some effect on the nonresponse of the SLC, because in that study telephone interview was not allowed. Fourthly, the 1994 survey was the first large survey at Statistics Finland conducted by using the CAPI-method, and the resources that were used to decrease nonresponse in the 1986 survey were now directed to giving technical ADP support to the interviewers. Negative attitudes towards surveys caused by the long lasting economic recession may also have decreased the willingness to participate.

Our analysis revealed that nonresponse was rather evenly distributed according to age and gender (Figure 1).

Because the survey tries to explore the living conditions, the well- and ill-being of the population in particular, claims for the representativeness in different social strata were high. Thus the essential task was to investigate how people in lower social strata are represented in the data compared with people in higher positions.

In Finland, we have many administrative data sources, which Statistics Finland is allowed to use for statistical purposes. Register data can be merged to survey data by the personal identification number (PIN). For example, in order to create the so-called "recession trap" data we had to include variables from files maintained by the National Research and Development Centre for Welfare and Health (STAKES). Those variables contained information on people whose family had had economic problems in 1992. It is supposed here that the "registered ill-being" in 1992 might have some effects on people's life also in 1994.

The variables used in this comparison were:

1) unemployment (some member of the family being unemployed in 1992),
2) indicator of being poor (the income of the family was less than half of the median income of the households),
3) indicator of being "overindebted" (household had to apply for various kinds of aid from social authorities because of their burden of loans)

4) indicator of a household receiving living allowance (living allowance is discretionary and a temporary form of aid paid to the family because of insufficient incomes).

A combined indicator called the "recession trap" variable, was also constructed. If the family had at least one of the defects mentioned above, the indicator was positive, otherwise 0.

Figure 2 shows that the above-mentioned economic ill-being indicators did not decrease the response rates among females to any considerable degree. In families which had experienced unemployment or were "overindebted" the response rate was on average even slightly higher.

The response rate of males, instead, decreased if the family belonged to the recession trap group. Unemployment is nowadays not as sensitive a social measure as the convential poverty measures (being poor or receiving living allowance). If the family of the male had received living allowance from the authorities, the response rate was 10 percentage points lower on average. The nonresponse of divorced males was also 8 percentage points higher than the average of all males, whereas the response rate of divorced females was about the average. This indicates that men in the lower social strata or uncertain social situation are underrepresented in the results. And it is also possible that among the sensitive groups, such as those receiving living allowance, people who are worst off have an even lower participation rate than those who are not that badly off (e.g. receive support only temporarily).

Fig. 1. Response rate in the 1994 SLC by age and by gender.

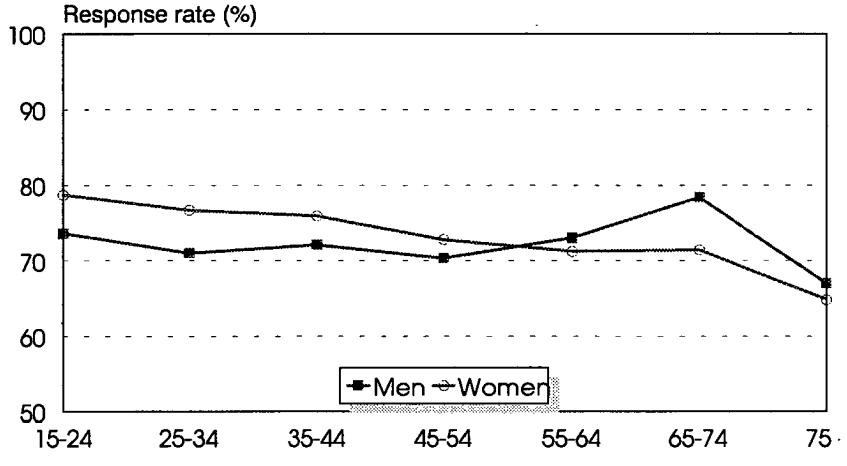
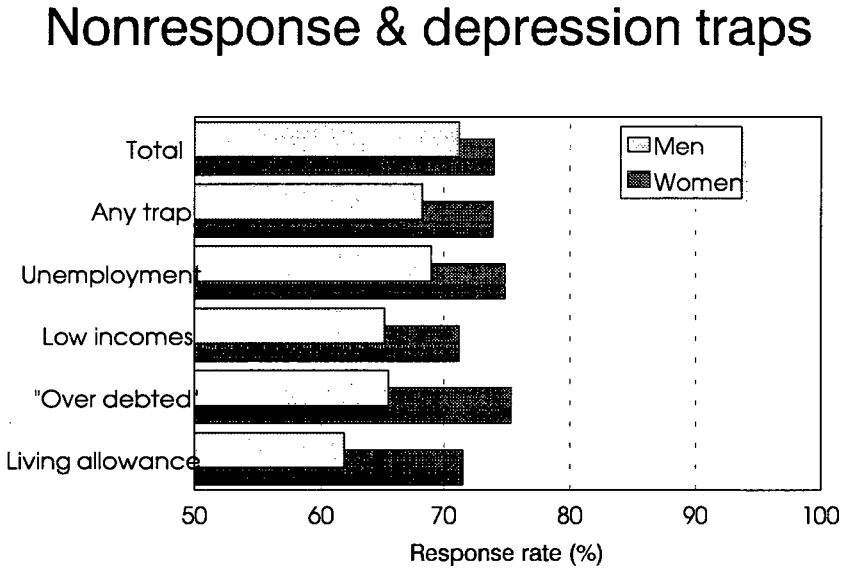


Fig. 2. Response rate of the 1994 SLC according to different ill-being criteria and gender.



THE 1987–1988 TIME USE SURVEY

Introduction

The Time Use Survey (TUS) is one of the large-scale surveys conducted on a fairly regular basis in a number of countries. The first Finnish TUS was carried out in the autumn of 1979, the second between April 1, 1987 – March 31, 1988. The next TUS is currently being developed and will be carried out in the late 1990's. All time use surveys are internationally comparable with respect to time use variables. This comparability is, however, not so clear from the point of view of survey designs, the dealing with nonresponse included. Naturally, there are several problems with nonresponse, in particular, if it is expected that the survey should provide reasonably accurate estimates of all the time use variables. This means that time use should be measured and estimated reliably over all seasons, all days of the week and all hours of the day. Furthermore, a survey should give the time use estimates for various socio-economic groups, for various members of households as well as for various age groups of people. Correspondingly, the sampling design should ensure that these different targets are sufficiently well taken into account.

Sampling design

The sampling design for the 1987–88 TUS was constructed in two phases. First, the essential aspects of the research objectives were taken into account, and a sample was formed to evenly cover the whole year. The first phase included the construction of appropriate pre-strata, from which a random sample was selected. The second phase involved the production of new pre-strata, numbering as many as 365. The original sample was drawn from the register-based sampling frame. It was only updated as far as the interview period is concerned. This is not the worst problem, however. The most critical problems are those concerning responding. There are problems due to (i) unit nonresponse, and (ii) response only if postponement is possible. Postponement is a commonly employed technique to increase response rates in diary surveys, but naturally it is subject to certain conditions, which must be determined in advance. It was required that the day of the week must be the same as the original day in the case of postponement in the Finnish TUS. The maximum time for postponement was three weeks.

Nonresponse and postponement

Figures 1 and 2 indicate that the option of postponement indeed increases the response rate a great deal, presuming every postponement would otherwise result in nonresponse. The main reason for postponement is that the respondent could not be contacted at his address for an interview due to vacation, work or respective reason. The main reason for unit nonresponse is refusal.

Fig. 1. Unit nonresponse and postponements of diary keeping by main reason in the Finnish 1987–88 Time Use Survey.

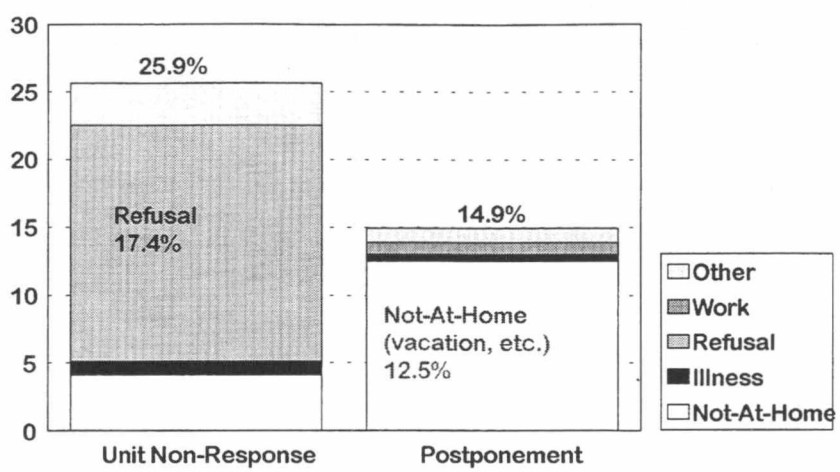
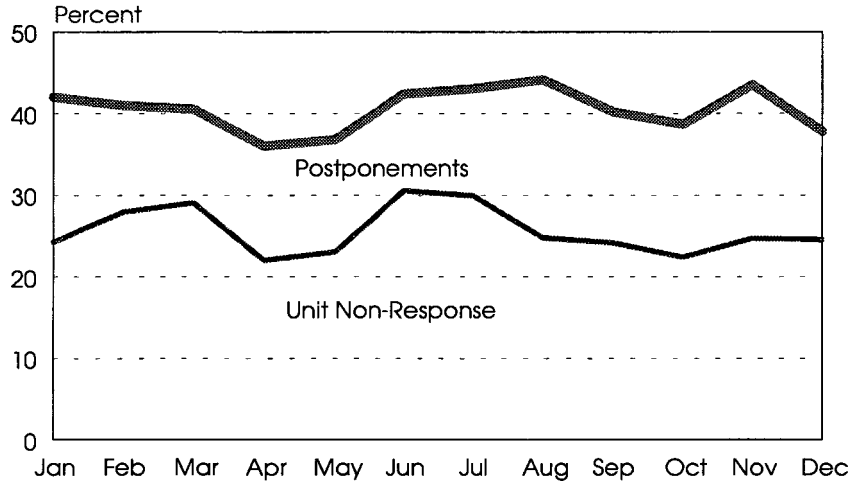


Fig. 2. Unit nonresponse and postponement of diary keeping by month in the Finnish 1987–88 Time Use Survey.



One might reckon that 'hard' refusals do not depend on the season to any greater degree, but rather on individual characteristics. In reality, some refusals are due to vacations, thus overestimating the number of refusals. There are noticeable differences between months, the nonresponse rates being highest during typical vacation months, such as June and July, but also in March due to Easter.

Postponements are also unevenly distributed over the year. The highest postponement rates are recorded for August, November and January, the point of time being determined according to the original interview period. The worst months for both nonresponse and postponing, cumulatively, were August, July, January, November and June (Fig.2.).

Seasonality is not the only important indicator of postponement or nonresponse. Laaksonen and Pääkkönen (1992) examined the background data of the individuals, using information obtained from registers, which naturally entailed certain restrictions. Logistic regression models were constructed to explain both the probability of response and that of postponement. The former model was considerably better than the latter one. For the response model they found five significant variables with 80 cells, whereas the model for postponement consisted of three significant variables with 35 cells.

The best explanatory variables for postponement were 'region' and 'educational level,' followed by 'season.' The best explanatory variables for response were 'region,' 'age,' 'gender,' 'educational level' and 'taxable income,' but also 'urban/rural' and 'season' were significant. The results indicate that middle-aged persons with a higher educational level responded best of all. Elderly females had the lowest response rate, but their record was

only slightly worse than that of the group of the middle-aged with low educational level. When other characteristics were taken into account, it was observed that high-income earners show close to perfect response. On the other hand, residence in the south and in urban areas affects response negatively, especially when diary records were to be kept in summer.

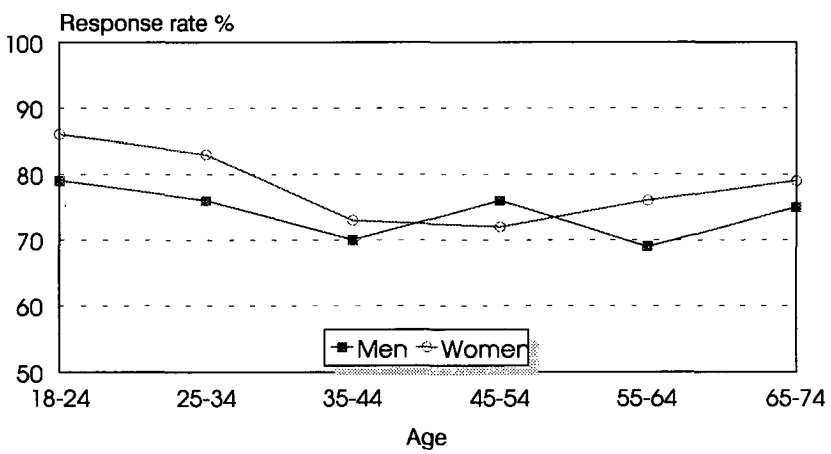
THE 1991/1992 SURVEY OF SEXUAL BEHAVIOUR

The survey was conducted by the Department of Sociology of the University of Helsinki. The funds for the survey were provided by the Finnish Academy. Statistics Finland took care of the fieldwork and the basic data analysis. The sample consisted of 3,049 persons aged 18–74 years. The interviews were conducted between November 1991– February 1992.

The overcoverage (deaths, emigrated, permanent illness) was somewhat larger than in Finnish population surveys in general, 85 persons (2.8 % of the original sample). Thus, it seems that some non-response cases were included in these figures. The net sample size was 2,964 persons.

2,252 interviews were completed and the response rate amounted to 76 %. A self-administered questionnaire was used in the middle of the interview to gather information on the most sensitive topics (e.g. number of sex partners, frequency of intercourse, homosexuality). 2,196 self-administered questionnaires were obtained (74 % of the net sample). The interview time averaged 78 minutes (interview + questionnaire).

Fig. 1. Response rates in the 1991/1992 Finnish Sex Survey.



The response rate was 78 % among men and 74 % among women. Especially women younger than 35 years participated actively in the survey. Figure 1 also shows that the response rate of women aged 35–55 was lower than in the other age groups. It could be related to the interviewer corps: practically all interviewers were women, their average age being 44 years.

Refusal was the main reason for nonresponse: 22 % of the net sample. However, only 3 % of the net sample gave the topic of the survey as the reason for refusal. 3 % of the sample were not contacted during the fieldwork time.

THE 1992 SURVEY OF DRINKING HABITS

A methodological study was connected to the 1992 Survey of Drinking Habits. The aim of this separate study was to compare the results of two different interviewer organisations: Statistics Finland and the Foundation for Alcohol Studies in Finland (FFAS). The interviewers of the latter were men and of the former, women. The target population consisted of the resident population aged from 15 to 69 years. Most general guidelines were kept as similar as possible in the two organisations. However, each conducted the actual fieldwork in its own way.

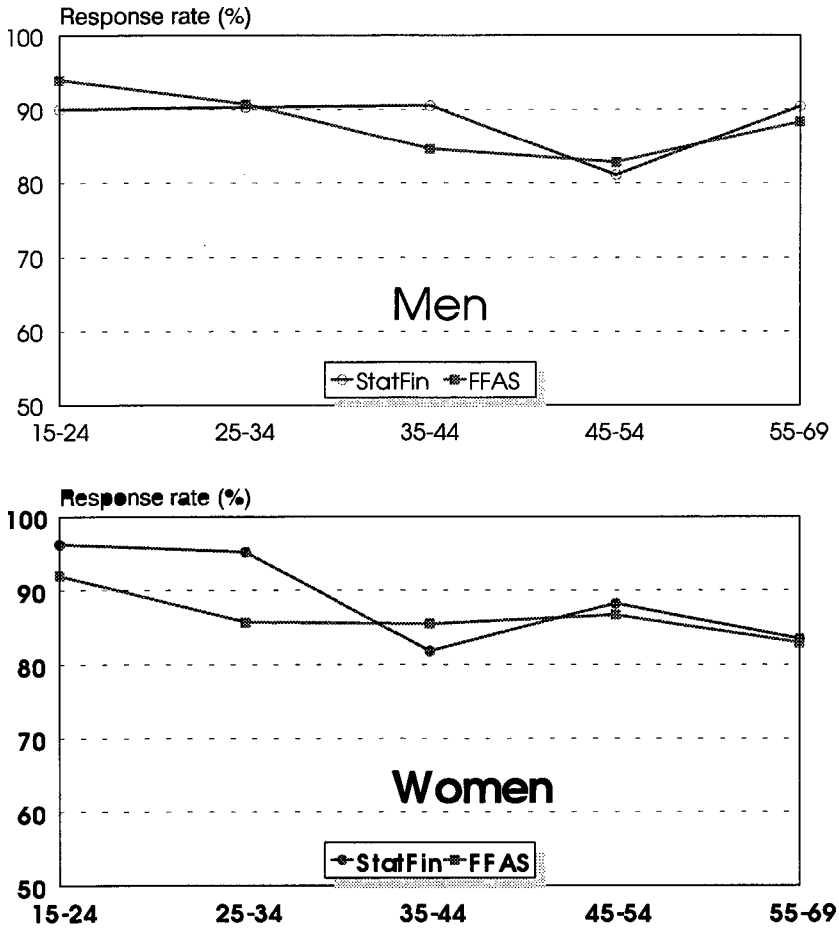
The response rates of both organisations were higher than those normally achieved in Statistics Finland's surveys on individuals (face-to-face interviews). The response rate of the FFAS has exceeded 90 per cent in its previous surveys. One reason for this traditionally high response rate may be the fact that alcohol is an interesting survey topic in Finland. The fact that potential nonresponse cases were reassigned (e.g. changing the interviewer) and the number of contact attempts was not limited can be regarded as other contributory reasons. On the other hand, the fieldwork period was short lasting only one month, and no incentives were given.

Table 1. Nonresponse rates.

Organisation	Response rate* (%)		Refusals*	Sample size*	Overcoverage
	Total	Men	Total (%)	(n) Total	(%)
Statistics Finland	88,7	88,8	8,4	936	1,5
FFAS	87,2	87,9	7,7	3,954	0,6

* overcoverage excluded

Figure 1. Response rates of the two interview organisations by sex.



Male interviewers (FFAS) had a slightly higher response rate when interviewing young men. Correspondingly female interviewers obtained the best results as regards young women.

VICTIMISATION SURVEYS

Violence is regarded as a more severe problem in Finland than in most other Western European countries, although it has not increased recently. The number of property crimes, instead, has increased. These are results of victimisation surveys conducted by Statistics Finland.

The older **victimisation survey** tradition builds on the OECD recommendations of measuring physical safety. It includes the criminal victimisation, the fear of being victimised, and the victimisation of accidents (traffic, work, home, sports, other leisure time). Those surveys were conducted as an addendum to the Finnish Labour Force Survey. The average interview time (LFS + victimisation survey) was 20 minutes.

The response rate in the victimisation surveys has decreased. However, the decrease between the 1980 and 1988 survey depended mostly on the different sampling frame and changes in fieldwork methods. Nonresponse was increased by the practice that people living in institutions and those having no permanent residence were included in the sampling frame in the 1988 survey, whereas in the 1980 survey they were excluded. The practice of the LFS to accept proxy interviews (questions of the LFS were asked of some other person living in the household than the sampled person) also influenced nonresponse because proxy respondents were not included in the victimisation survey. The increase in refusals was negligible.

Year of the study	Response rate* (%)	Refusals (%)	Non contact (%)	Sample size (n)
1980	92.2	3.9	2.2	10,405
1988	87.0	4.0	6.0	14,861
1993	85.3	4.4	6.2	5,010

*Overcoverage excluded

The international crime survey (ICS) has been conducted in 1989 and 1992. Many countries have participated twice, some only once. Sample sizes of the ICS in different countries are small: 1 000 – 2 000 persons aged 16 and over. In most countries the interview is conducted by using the CATI method, i. e. only persons who can be contacted on the telephone are interviewed.

In 1992 the response rate amounted to 84 % in Finland, which is 14 percentage points higher than in 1989. The response rate was calculated by using persons who have a telephone in their household (82 % of the original sample). One reason for the increase in the response rate was probably the change of interviewer organisation between the 1989 and the 1992 studies. The former was conducted by a commercial survey organisation, the latter by Statistics Finland.

Response rates differ greatly from one country to another, i.e. from 30 to 85 per cent. Such differences certainly have an impact on data quality. Thus it is rather difficult to make international comparisons between the crime survey results.

NONRESPONSE IN THE BUSINESS AND INCOME STATISTICS OF FARMING

The Business and Income Statistics of Farming published yearly by Statistics Finland, are based on sample surveys where the data are compiled from the tax data of farmers. A special statistical questionnaire is used to solicit additional information concerning income on a more detailed level. Statistics include information on incomes, expenditures and subsidies of farming by production sector. The sampling design is a stratified simple random sampling and the sample size is about

15 000 farms. The Farm Register serves as the sampling frame in which variables such as region, production sector and arable land are available for stratification. The Farm Register comprises all farms with at least two hectares of arable land under cultivation and a farm can be used as a sampling unit.

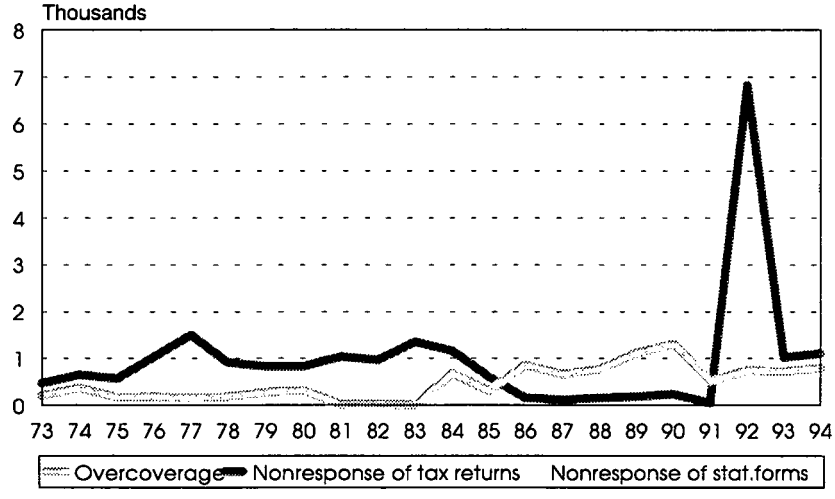
Fig. 1. Developments in the population, sample sizes and respondents in 1973 –1974.

The farms belonging to the population

Sample and respondents

The survey strategy comprises a rotating panel. The same sample was used both in 1973 and 1974, whereas during the period 1975–1983 one fifth of the sample changed yearly and since 1984 one third. Estimation is based on weighting in strata (Väisänen, 1995). The data were collected from the taxation register and tax returns up to 1991. In 1992 changes in taxation took place and the use of statistical forms was introduced into data collection. During the first year only about a half of the statistical forms were returned. The high nonresponse rate was due to the fact that the statistical forms were mailed directly to farmers from Statistics Finland and answering was voluntary and not connected to tax returns. The questionnaire of 1992 was rather complicated which was one reason for the high nonresponse. In 1993 and 1994 the data were gathered by using both tax returns received as copies from the tax offices and additional statistical forms. Finland joined the European Union in 1995 and farmers had to fill in several new questionnaires and forms in order to seek subsidies in the same period when the survey data were collected. This explains the higher nonresponse compared to the year before.

Fig. 2. Nonresponse and overcoverage.



Nonresponse in statistical forms was higher than nonresponse in tax forms. Filing of tax returns is obligatory for all farms liable to pay taxes. Nonresponse in tax returns consisted of cases where the owner had another holding in some other municipality of permanent residence and the tax returns were filed in that municipality. Tax returns were missing also from those farms which had made an appeal in tax cases. Hot deck imputation based on the data of tax returns was used for the item nonresponse of the statistical forms.

Table 1. Sample and nonresponse in the Business and Income Statistics of Farming.

Year	Population	Res-ponses	Over-coverage	Nonr.of tax returns	Non-response rate %	Nonr. of stat. form
73	239069	12156	218	472	3.74	–
74	225731	11833	370	643	5.15	–
75	215900	14485	172	569	3.78	–
76	203903	13443	184	1037	7.16	–
77	197620	14932	161	1501	9.13	–
78	190881	15423	184	907	5.55	–
79	185813	14152	284	834	5.57	–
80	181737	15292	322	829	5.14	–
81	180660	14829	26	1027	6.48	–
82	176207	15429	27	962	5.87	–
83	170534	15138	0	1353	8.20	–
84	159349	14613	674	1160	7.35	–
85	158627	15558	308	606	3.75	–
86	148615	15034	858	154	1.01	–
87	146576	15194	661	111	0.73	–
88	141341	15084	774	155	1.02	–
89	132646	14438	1119	180	1.23	–
90	121032	14774	1308	242	1.61	–
91	122296	14096	515	54	0.38	–
92	119055	7675	743	6819	47.05	6819
93	114739	12889	713	1025	7.37	4015
94	105571	12232	807	1099	8.24	4649

LITERATURE

Djerf, K. (1991a). Some aspects of the Finnish household expenditure survey 1990. *Paper presented to the Workshop on Diary Surveys, February 18–20, Stockholm, Sweden.* 15 p.

Djerf, K. (1991b). Survey-tutkimuksissa käytetyistä otanta- ja tiedonkeruumenetelmistä, teoksessa *Tilastokeskuksen henkilösurveyt nyt ja tulevaisuudessa.* Raportti 10.6.1991 pidetystä seminaarista. Tilastokeskus, Raportti nro 1991:4, 14–23. [Sampling and data collection procedures in the Surveys conducted by the Statistics Finland, in the *Statistics Finland Surveys on Persons and Households now and in the future.* Report from the seminar 10.6.1991. Statistics Finland, Report no 1991:4, 14–23. (in Finnish).]

Djerf, K. (1991c). Cumulating the nonresponse in four phases of data collection of the Household expenditure survey 1990. *Paper presented to the Second International Workshop on Household Survey Nonresponse, October 16–18, Washington DC, USA.* 19 p.

Djerf, K. (1993). Sampling, nonresponse and estimation of results in the 1991 leisure survey, in *Culture of the Everyday*, eds. Mirja Liikkanen and Hannu Pääkkönen. Statistics Finland, Culture and the media 1994:3. Helsinki, 128–131.

Djerf, K. (1994a). Vuoden 1992 juomatapatutkimuksen otanta-asetelma ja estimointi. *Suomen Tilastoseuran vuosikirja* 1993, 205–220. Helsinki. [Sampling Design and Estimation of 1992 Drinking Habits Survey (In Finnish with English abstract), *The Yearbook of the Finnish Statistical Society* 1993, 205–220. Helsinki]

Djerf, K. (1994b). Effects of Post-stratification on the Estimates of the Finnish Labour Force Survey. 12 p. *Manuscript submitted to the Journal of Official Statistics.*

Djerf, K. (1994c). Yleistetty raking-menetelmä otantatutkimuksen estimoinnin tehostamismenetelmänä. Sovellus työvoimatutkimuksen aineistoon. Jyväskylän yliopisto, tilastotieteen laudaturtyö, syyskuu 1994. 38 s. + 5 s. liitteitä. [Generalized raking – a method to utilize auxiliary information in estimation *University of Jyväskylä, masters' thesis in statistics*, September 1994, 38 p + appendices.]

Djerf, K. (1995a). Otanta-asetelma ja painotus, teoksessa *Elinolotutkimus 1994. Aineiston keruu.* toim. Markku Heiskanen. Tilastokeskus, Muistio 1995:2, 20–31. [*Sampling Design and Weighting, in Survey of Living Conditions 1994. Data Collection.* Ed. Markku Heiskanen. Statistics Finland, Memorandum 1995:2, 20–31.]

Djerf, K. (1995b). Effects of Calibration in two Finnish Surveys. *Proceedings of the Nordic Meeting of Statisticians, August 2–5, 1995, Rebild, Denmark. Nordic statistisk sekretariat*, 746–756.

Djerf, K. and Lindqvist, M. (1993a). Kotitaloustiedustelun tutkimusmenetelmä ja tulosten luotettavuus, teoksessa *Kotitalouksien kulutusmenot 1990.* Tilastokeskus, Tulot ja kulutus 1992:17, Helsinki, 57–58.[Description of methods and quality of the Household Budget Survey 1990 (in Finnish), in *Household expenditures 1990.* Statistics Finland, Income and consumption 1992:17. Helsinki, 57–58.

Djerf, K. and Lindqvist, M. (1993b). *Kotitaloustiedustelun laatuselvitys.* Tilastokeskus, Keskustelumuistioita 3. Helsinki. 46 s. + liitteitä. [*Quality of the Household Budget Survey 1990* (in Finnish)]. Statistics Finland: Working Papers 3. Helsinki: 46 p. + appendices.

Djerf, K. and Väisänen, P. (1993). Effects of Post-stratification on the Estimates of the Finnish Labour Force Survey. *Bulletin of the International Statistical Institute, 49th session, Contributed Papers, Book 1.* Firenze, 375–376.

Ekholm, A. and Laaksonen, S. (1991). Weighting via Response Modelling in the Finnish Household Budget Survey. *Journal of Official Statistics*, 7,2, 325–337.

Household Budget Survey 1976. Standard Errors of Consumption Expenditure (Vol. IV). (1980), Central Statistical Office of Finland,Tilastollisia tiedonantoja 62. Helsinki. (in Finnish and Swedish)

Household Budget Survey 1981. Quality Report (Vol. III). Central Statistical Office of Finland, Statistical Studies 71 (1986), Statistics Finland, Helsinki. (in Finnish and Swedish)

Household Budget Survey 1981, User's Guide. (1985) Central Statistical Office of Finland, Muisio 100, Helsinki. (in Finnish)

Household Budget Survey 1985. Quality Report. Central Statistical Office of Finland, Income and Consumption 1988:5, Statistics Finland, Helsinki. (in Finnish with English Summary)

Household Budget Survey 1994. Kotitaloustiedustelun tutkimusmenetelmä ja tulosten luotettavuus, teoksessa Markku Lindqvist ja Pertti Kangassalo (toim.) *Kotitalouksien kulutusmenot 1994.* Tilastokeskus, Tulot ja kulutus 1996:4, 39–40.[Description of methods and quality of the Household Budget Survey 1994, in Markku Lindqvist ja Pertti Kangassalo (eds.) *Household Consumption Expenditure 1994*, Statistics Finland, Income and consumption 1996:4, 39–40. (in Finnish).

Laaksonen, S. (1988). *Katovirheen korjaus kotitalousaineistossa.* English Summary: *Correcting for Nonresponse in Household Data* (10 pages). Central Statistical Office of Finland. Research Reports 147. 109 pages.

Laaksonen, S. (1991a). Adjustment for Non-response in Two-year Panel Data: Applications to Problems of Household Income Distribution. *The Statistician* 40, 153–168.

Laaksonen, S. (1991b). *Comparative Adjustments for Missingness in Short-term Panels. Applications to Questions of Household Income Distribution.* Central Statistical Office, Research Reports 179. 66 pages.

Laaksonen, S. (1991c). Talletus- ja pörssiosaketietojen imputointi vuosien 1987–88 säästämis- ja velkaantumistutkimuksessa (Imputation of missing values of deposits and bonds in the Finnish two-year panel survey 1987–1988). In: Ahlqvist, K., Kangassalo, P., Laaksonen, S. and Säylä, M. Raportti 1991.5, Statistics Finland, Section 6, 44– 54.

Laaksonen, S. (1991d). A General Strategy for Handling Household Survey Nonresponse Data in Practical Situations. *48th Session of the ISI, Cairo. Contributed Paper*, 401–402.

Laaksonen, S. (1992a). Adjustment Methods for Non-response and their Application to Finnish Income Data. *Statistical Journal of Economic Commission for Europe of United Nations 9*, 125–137.

Laaksonen, S. (1992b). *Handling Household Survey Nonresponse Data.* The Finnish Statistical Society. Research Reports 13.

Laaksonen, S. and Pääkkönen, H. (1992). Some Methodological Aspects on the Use of Time Budget Survey Data. In: Kirjavainen et al (1992). *Housework Time in Bulgaria and Finland.* Statistics Finland. Research Reports 193, 86–104.

Nonresponse Research of the 1976 Household Budget Survey (1979). Central Statistical Office of Finland, Muistio 52, Helsinki. (in Finnish)

Reporting the Quality of Statistics. Central Statistical Office of Finland, Handbook 23, Helsinki. (in Finnish)

Statistics Finland. Unpublished statistical references. Helsinki. (in Finnish)

Väisänen, P. (1995). Sample Design in the Finnish Agricultural Income Survey. *Proceedings of the 1995 Kansas State University Conference on Applied Statistics in Agriculture. April 23–25, 1995. Department of Statistics, Kansas State University, Manhattan, Kansas.*

The workshop participants of the second last session row by row (from left to right), Friday 27 October:
1: Lars Lyberg Sweden, Lilli Japec Sweden, Bob Groves USA, Edith de Leeuw Netherlands, Joop Hox Nethertlands, Mick Couper USA/South Africa,
2: Stephen Mack USA, Kari Djerf Finland, Pamela Campanelli UK, Markku Heiskanen Finland, Seppo Laaksonen Finland, Markku Lindqvist Finland,
3: Vasja Vehovar Slovenia, Metka Zaletel Slovenia, Trena Ezzati-Rice USA, Sylvie Michaud Canada, Frank Mayda Canada, Terry O'Connor USA, Paavo Väisänen Finland, Erkki Pahkinen Finland,
4: Raoul Depoutot Eurostat/Luxembourg/France, Rita Petroni USA, Wim de Heer Netherlands, Martin Luppes Netherlands, Jan van den Brakel Netherlands, Peter Hauptmanns Germany,
5: Bo Möller Denmark, Marina Signore Italy, Marriette Vosmer Netherlands, Kate Foster UK, John King UK, Risto Lehtonen Finland, Pertti Kangassalo Finland,
6: Hannu Pääkkönen Finland, Leena Hietaniemi Finland, Irja Blomqvist Finland, Iiris Niemi Finland, Vesa Kuusela Finland.

The presence rate was 90 per cent since the following persons were absent at the moment the photo was taken:
Preston Jay Waite USA, Marja Sjöblom Finland, Adam Marton Hungary, Terry O'Connor USA and Carl-Erik Särndal Canada/Sweden/Finland.

TUTKIMUKSIA-SARJA

Tilastokeskus on julkaissut Tutkimuksia v. 1966 alkaen, v. 1990 lähtien ovat ilmestyneet seuraavat:

164. **Henry Takala,** Kunnat ja kuntainliitot kansantalouden tilipdossa. Tammikuu 1990. 60 s.

165. **Jarmo Hyrkkö,** Palkansaajien ansiotasoindeksi 1985=100. Tammikuu 1990. 66 s.

166. **Pekka Rytkönen,** Siivouspalvelu, ympäristöhuolto ja pesulapalvelu 1980-luvulla. Tammikuu 1990. 70 s.

167. **Jukka Muukkonen,** Luonnonvaratilinpito kestävän kehityksen kuvaajana. 1990. 119 s.

168. **Juha-Pekka Ollila,** Tieliikenteen tavarankuljetus 1980-luvulla. Helmikuu 1990. 45 s.

169. **Tuovi Allén – Seppo Laaksonen – Päivi Keinänen – Seija Ilmakunnas,** Palkkaa työstä ja sukupuolesta. Huhtikuu 1990. 90 s.

170. **Ari Tyrkkö,** Asuinolotiedot väestölaskennassa ja kotitaloustiedustelussa. Huhtikuu 1990. 63 s.

171. **Hannu Isoaho – Osmo Kivinen – Risto Rinne,** Nuorten koulutus ja kotitausta. Toukokuu 1990. 115 s.

171b. **Hannu Isoaho – Osmo Kivinen – Risto Rinne,** Education and the family background of the young in Finland. 1990. 115 pp.

172. **Tapani Valkonen – Tuija Martelin – Arja Rimpelä,** Eriarvoisuus kuoleman edessä. Sosioekonomiset kuolleisuuserot Suomessa 1971–85. Kesäkuu 1990. 145 s.

173. **Jukka Muukkonen,** Sustainable development and natural resource accounting. August 1990. 96 pp.

174. **Iiris Niemi – Hannu Pääkkönen,** Time use changes in Finland in the 1980s. August 1990. 118 pp.

175. **Väinö Kannisto,** Mortality of the elderly in late 19th and early 20th century Finland. August 1990. 50 pp.

176. **Tapani Valkonen – Tuija Martelin – Arja Rimpelä,** Socio-economic mortality differences in Finland 1971–85. December 1990. 108 pp.

177. **Jaana Lähteenmaa – Lasse Siurala,** Nuoret ja muutos. Tammikuu 1991. 211 s.

178. **Tuomo Martikainen – Risto Yrjönen,** Vaalit, puolueet ja yhteiskunnan muutos. Maaliskuu 1991. 120 s.

179. **Seppo Laaksonen,** Comparative Adjustments for Missingness in Short-term Panels. April 1991. 74 pp.

180. **Ágnes Babarczy – István Harcsa – Hannu Pääkkönen,** Time use trends in Finland and in Hungary. April 1991. 72 pp.

181. **Timo Matala,** Asumisen tuki 1988. Kesäkuu 1991. 64 s.

182. **Iiris Niemi – Parsla Eglite – Algimantas Mitrikas – V.D. Patrushev – Hannu Pääkkönen,** Time Use in Finland, Latvia, Lithuania and Russia. July 1991. 80 pp.

183. **Iiris Niemi – Hannu Pääkkönen,** Vuotuinen ajankäyttö. Joulukuu 1992. 83 s.

183b. **Iiris Niemi – Hannu Pääkkönen – Veli Rajaniemi – Seppo Laaksonen – Jarmo Lauri,** Vuotuinen ajankäyttö. Ajankäyttötutkimuksen 1987–88 taulukot. Elokuu 1991. 116 s.

184. **Ari Leppälahti – Mikael Åkerblom,** Industrial Innovation in Finland. August 1991. 82 pp.

185. **Maarit Säynevirta,** Indeksiteoria ja ansiotasoindeksi. Lokakuu 1991. 95 s.

186. **Ari Tyrkkö,** Ahtaasti asuvat. Syyskuu 1991. 134 s.

187. **Tuomo Martikainen – Risto Yrjönen,** Voting, parties and social change in Finland. October 1991. 108 pp.

188. **Timo Kolu,** Työelämän laatu 1977–1990. Työn ja hyvinvoinnin koettuja muutoksia. Tammikuu 1992. 194 s.

189. **Anna-Maija Lehto,** Työelämän laatu ja tasa-arvo. Tammikuu 1992. 196 s.

190. **Tuovi Allén – Päivi Keinänen – Seppo Laaksonen – Seija Ilmakunnas,** Wage from Work and Gender. A Study on Wage Differentials in Finland in 1985. 88 pp.

191. **Kirsti Ahlqvist,** Kodinomistajaksi velalla. Maaliskuu 1992. 98 s.

192. **Matti Simpanen – Irja Blomqvist,** Aikuiskoulutukseen osallistuminen. Aikuiskoulutustutkimus 1990. Toukokuu 1992. 135 s.

193. **Leena M. Kirjavainen – Bistra Anachkova – Seppo Laaksonen – Iiris Niemi – Hannu Pääkkönen – Zahari Staikov,** Housework Time in Bulgaria and Finland. June 1992. 131 pp.

194. **Pekka Haapala – Seppo Kouvonen,** Kuntasektorin työvoimakustannukset. Kesäkuu 1992. 70 s.

195. **Pirkko Aulin-Ahmavaara,** The Productivity of a Nation. November 1992. 72 pp.

196. **Tuula Melkas,** Valtion ja markkinoiden tuolla puolen. Joulukuu 1992. 150 s.

197. **Fjalar Finnäs,** Formation of unions and families in Finnish cohorts born 1938–67. April 1993. 58 pp.

198. **Antti Siikanen – Ari Tyrkkö,** Koti – Talous – Asuntomarkkinat. Kesäkuu 1993. 167 s.

199. **Timo Matala,** Asumisen tuki ja aravavuokralaiset. Kesäkuu 1993. 84 s.

200. **Arja Kinnunen,** Kuluttajahintaindeksi 1990=100. Menetelmät ja käytäntö. Elokuu 1993. 89 s.

201. **Matti Simpanen,** Aikuiskoulutus ja työelämä. Aikuiskoulutustutkimus 1990. Syyskuu 1993. 150 s.

202. **Martti Puohiniemi,** Suomalaisten arvot ja tulevaisuus. Lokakuu 1993. 100 s.

203. **Juha Kivinen – Ari Mäkinen,** Suomen elintarvike- ja metallituoteteollisuuden rakenteen, kannattavuuden ja suhdannevaihteluiden yhteys; ekonometrinen analyysi vuosilta 1974 –1990. Marraskuu 1993. 92 s.

204. **Juha Nurmela,** Kotitalouksien energian kokonaiskulutus 1990. Marraskuu 1993. 108 s.

205a. **Georg Luther,** Suomen tilastotoimen historia vuoteen 1970. Joulukuu 1993. 382 s.

205b. **Georg Luther,** Statistikens historia i Finland till 1970. December 1993. 380 s.

206. **Riitta Harala – Eva Hänninen-Salmelin – Kaisa Kauppinen-Toropainen – Päivi Keinänen – Tuulikki Petäjäniemi – Sinikka Vanhala,** Naiset huipulla. Huhtikuu 1994. 64 s.

207. **Wangqiu Song,** Hedoninen regressioanalyysi kuluttajahintaindeksissä. Huhtikuu 1994. 100 s.

208. **Anne Koponen,** Työolot ja ammattillinen aikuiskoulutus 1990. Toukokuu 1994. 118 s.

209. **Fjalar Finnäs,** Language Shifts and Migration. May 1994. 37 pp.

210. **Erkki Pahkinen – Veijo Ritola,** Suhdannekäänne ja taloudelliset aikasarjat. Kesäkuu 1994. 200 s.

211. **Riitta Harala – Eva Hänninen-Salmelin – Kaisa Kauppinen-Toropainen – Päivi Keinänen – Tuulikki Petäjäniemi – Sinikka Vanhala,** Women at the Top. July 1994. 66 pp.

212. **Olavi Lehtoranta,** Teollisuuden tuottavuuskehityksen mittaminen toimialatasolla. Tammikuu 1995. 73 s.

213. **Kristiina Manderbacka,** Terveydentilan mittarit. Syyskuu 1995. 121 s.

214. **Andres Vikat,** Perheellistyminen Virossa ja Suomessa. Joulukuu 1995. 52 s.

215. **Mika Maliranta,** Suomen tehdasteollisuuden tuottavuus. Helmikuu 1996. 189 s.

216. **Juha Nurmela,** Kotitaloudet ja energia vuonna 2015. Huhtikuu 1996. 285 s.

217. **Rauno Sairinen,** Suomalaiset ja ympäristöpolitiikka. Elokuu 1996. 179 s.

218. **Johanna Moisander,** Attitudes and Ecologically Responsible Consumption. August 1996. 159 pp.

219. **Laaksonen Seppo (ed.),** International Perspectives on Nonresponse. Proceedings of the Sixth International Workshop on Household Survey Nonresponse. December 1996. 240 pp.

220. **Jukka Hoffrén,** Metsien ekologisen laadun mittaaminen. Elokuu 1996. 79 s.

221. **Jarmo Rusanen – Arvo Naukkarinen – Alfred Colpaert – Toivo Muilu.** Differences in the Spatial Structure of the Population Between Finland and Sweden in 1995 – A GIS Viewpoint. December 1996.

222. **Anna-Maija Lehto.** Työolot tutkimuskohteena. Marraskuu 1996. 289 s.

The Research Report Series describe the Finnish society in the light of up-to-date research results. Scientific studies that are carried out at Statistics Finland or that are based on the data sets of Statistics Finland are published in the series. The series also includes collections of scientific articles such as conference proceedings.

These proceedings consist of the 20 scientific papers submitted to the International Workshop on Household Survey Nonresponse held in Helsinki in 1995. Articles consider both procedures to reduce nonresponse at the data collection stage and post-survey adjustments, i.e. statistical methods for adjusting bias caused by nonresponse and coverage errors.

The volume serves as a timely handbook on current statistical methodology for handling nonresponse in empirical studies intended for researchers, students and survey statisticians.

MYYNTI
Tilastokeskus
Myyntipalvelu
PL 3B
00022 TILASTOKESKUS
Puh (09) 1734 2011
Fax (09) 1734 2474

FÖRSÄLJNING
Statistikcentralen
Försäljningstjänsten
PB 3B
00022 STATISTIKCENTRALEN
Tfn (09) 1734 2011
Fax (09) 1734 2474

ORDERS
Statistics Finland
Sales Services
P.O. Box 3B
00022 STATISTICS FINLAND
Tel. + 358 9 1734 2011
Fax + 358 9 1734 2474

E-mail: tilastomyynti@stat.fi

ISSN 0355-2071
ISBN 951-727-217-0

31