Journal of Official Statistics
Vol. 1, No. 3, 1985, pp. 287–300

On the Treatment of Nonresponse in Sample Surveys

J. G. Bethlehem and H. M. P. Kersten[1]

Abstract: In carrying out a sample survey one is often faced with the problem of nonresponse and the possibility of bias which may be induced by this phenomenon. In this paper some techniques are presented to deal with nonresponse. A criterion is proposed to judge whether the possible effect of nonresponse is unacceptably high. Additionally, an inexpensive technique, called the Basic Question Procedure, is suggested to obtain insight and to reduce a possible bias due to nonresponse. Finally, a weighting technique based on linear models is discussed.

Key words: Nonresponse; adjustment procedures; bias.

1. Introduction

Nonresponse is the phenomenon that units in the selected sample and eligible for the sample survey do not provide the requested information, or that the provided information is unusable. The situation in which all requested information on a unit is missing is called unit nonresponse. If information is missing on some items only, we talk of item nonresponse. We will restrict ourselves to the case of unit nonresponse. Due to nonresponse the sample size is smaller than expected and, especially in case of high nonresponse rates, may give rise to increased variances. The main problem caused by nonresponse is that estimates of population characteristics may be biased. This situation occurs if due to nonresponse, some groups in the population are over- or underrepresented and these groups behave differently with respect to the characteristics to be investigated.

Indeed, estimators must be assumed biased unless very convincing evidence to the contrary is provided. A follow-up study of the Dutch Victimization Survey showed that people who are afraid when they are alone at night are less inclined to participate in the survey. In the Dutch Housing Demand Surveys it turned out that people who refused, demand less housing than people who responded. For the Survey on the Mobility of the Dutch Population, it is obvious that mobile people are underrepresented among the respondents.

Recently, the magnitude of the nonresponse in many Dutch sample surveys is increasing to such an extent that without special adjustment techniques, one has to reckon with a decrease in the quality of the results. Table 1 presents nonresponse rates for a number of Dutch surveys which are carried out by the Netherlands Central Bureau of Statistics. Also, the nonresponse rates due to refusals are presented.

[1] Department for Statistical Methods, P. O. Box 959, 2270 AZ Voorburg, The Netherlands. The views expressed in this paper are those of the authors and do not necessarily reflect the policies of the Netherlands Central Bureau of Statistics. The authors wish to express their gratitude to three anonymous referees for their valuable remarks on an earlier draft.

Table 1. Nonresponse Rates (NR) and Refusal Rates (RR) of some Dutch Sample Surveys in Percent

Year	Sample survey									
	Labour Force Survey		Survey of Consumer Sentiments		Survey on Well-Being of the Population		Survey on Mobility		Holiday Survey	
	NR	RR	NR	RR	NR	RR	NR	RR	NR	RR
1972			29	..						
1973	13	..	23	..						
1974			25	..	28	16				
1975	16	9	22	18					14	..
1976			28	19	23[1]	16			13	..
1977	13	7	31	21	30	17			19	9
1978			36	24			33	26	22	13
1979	19	..	37	24	35[2]	21	31	24	26	16
1980			39	25	39	24	32	23	26	..
1981	19	11	35	24			32	24	26	..
1982			40	27	36[1]	25	35	26	29	..
1983	18	..	37	24	42	26	34	27	26	17
1984									30	..

[1] elderly only. [2] young only. ..=not available.

It is difficult to compare response rates from different surveys, but for each of the surveys one can study nonresponse figures over a number of years. The magnitude of the nonresponse is determined by a large number of factors, including the subject of the survey, the target population, the time period, the length of the questionnaire, the quality of the interviewers, the fieldwork in general, etc. It is clear from Table 1 that nonresponse is a considerable problem. It has an impact on the cost of a survey since it takes more and more effort to obtain estimates with the same accuracy as originally specified in the sampling design.

Nonrespondents can be classified in three important groups. People in the first group refuse to cooperate. Sometimes it is possible to make an appointment for an interview at some later date. However, frequently the refusal can be final. Possible causes are fear of privacy intrusion and interview fatigue.

The second group of nonrespondents are the not-contactables. No contact is made due to that people are not at home, due to removal or due to other circumstances such as watchdogs, dangerous neighbourhoods or houses which are difficult to reach. Generally speaking, people are increasingly hard to contact. Important factors are smaller family sizes, greater mobility and a larger amount of spare time which is spent outdoors.

The third group of nonrespondents consists of people who are physically or mentally not able to cooperate during the fieldwork period. Also, language problems can cause this type of nonresponse.

To be able to build a well-founded theory of nonresponse adjustment, it is necessary to incorporate the phenomenon of nonresponse in general sampling theory. To that end, in Section 2, the Fixed Response Model is introduced. It then becomes clear how nonresponse affects estimators of population characteristics.

Section 3 proposes a criterion to judge whether the possible nonresponse bias is so high that additional measures become necessary. This criterion was applied in the 1981 Dutch Labour Force Survey.

A good way to obtain more information about nonrespondents is to select a sample from the nonrespondents and try the interview again. Particularly when people are not at home after a number of calls, this can be a successful means to obtain missing information. The technique can also be tried with refusals, but in that case it will take specially trained interviewers. Sampling the nonresponse is a costly and time-consuming process and therefore the technique is not applied very often. The Basic Question Procedure is a simpler and less costly procedure, which can be used for refusals but also for not-at-homes. If a selected person refuses to cooperate and there is no chance of obtaining a complete interview at another time, one can try to obtain answers on only one or two important questions of the survey. Use of the Basic Question Procedure need not be restricted to refusals. For not-at-homes and also other nonresponse categories these questions can possibly be answered afterwards by telephone or by mail. In these cases, attention should be paid to response effects. In this way information is collected about the possible biases of different nonresponse categories with regard to the most important topic. This technique is discussed in Section 4.

One of the most important and frequently used weighting techniques is poststratification. Using information about the target population, the sample is divided in strata and all respondents within a stratum are assigned the same weight. Application of poststratification can be difficult if the sample size is small compared with the number of strata, or if there is not enough population information. Recently at the Netherlands Central Bureau of Statistics, a more general theory of weighting based on linear models was developed. This new technique is discussed in Section 5 and illustrated by an example.

2. Some Theory on Nonresponse

Let the target population of the survey consist of N identifiable elements, which are labeled $1,2,\ldots,N$. Associated with each element k is an unknown value Y_k of a characteristic to be investigated. The objective of the sample survey is assumed to be estimation of the population mean

$$\bar{Y}=\frac{1}{N}\sum_{k=1}^{N} Y_k \ . \tag{2.1}$$

To estimate this quantity, a sample of size n is selected. The characteristics of the sampled elements are denoted by $y_1, y_2,\ldots,y_n$ (the y_i's are a subset of the Y_k's). If there is no nonresponse in this simple random sample, then the sample mean

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{n} y_i \ , \tag{2.2}$$

is an unbiased estimator of the population mean $\bar{Y}$.

In the literature, two different basic views are used of how nonresponse occurs. We label these views the Random Response Model and the Fixed Response Model. The Random Response Model assumes that every element in the population has a certain unknown probability of response. We restrict ourselves here to the Fixed Response Model. In this view, the population is assumed to consist of two mutually exclusive and exhaustive strata: the response stratum and the nonresponse stratum. Elements in the response stratum would participate in the survey with certainty, if selected in the sample, and elements in the nonresponse stratum would not participate with certainty, if selected. The Fixed Response Model can be regarded as a special case of the Random Response Model in which the response probabilities are either 0 or 1. The Fixed Response Model results in simple formulae which improve insight into the effects of nonresponse. Both views of nonresponse are discussed by Kalsbeek (1980) and Lindström et al. (1979). The Fixed Response Model is treated in Kersten and Bethlehem (1981) and the Random Response

Model is used in Bethlehem and Kersten (1981a, 1981b).

For the Fixed Response Model the size and contents of both strata are determined by a number of technical factors, such as the subject of the survey, the size and contents of the questionnaire and the time period of the fieldwork. Quantities which relate to the response stratum will be subscripted by r and quantities relating to the nonresponse stratum have the subscript nr. In this fashion, the number of elements of the response stratum is denoted by N_r and that of the nonresponse stratum by N_{nr}. Likewise the means of the characteristics in both strata are $\bar{Y}_r$ and $\bar{Y}_{nr}$. The difference

$$C = \bar{Y}_r - \bar{Y}_{nr} \tag{2.3}$$

between these two quantities is called the *contrast*, which is an indication of how respondents and nonrespondents differ on the average. The relative size of the nonresponse stratum is denoted by

$$Q = N_{nr}/N , \tag{2.4}$$

which is the expected value over all possible samples of the observed nonresponse rate n_{nr}/n.

If nonresponse occurs in a sample of size n, only observations on n_r respondents become available ($n=n_r+n_{nr}$). The mean $\bar{y}_r$ of these observations is an unbiased estimator of the mean $\bar{Y}_r$ of the response stratum, i.e.,

$$E(\bar{y}_r) = \bar{Y}_r . \tag{2.5}$$

If $\bar{Y}_r$ differs from $\bar{Y}$, then $\bar{y}_r$ is a biased estimator of $\bar{Y}$. The bias is equal to

$$B(\bar{y}_r) = E(\bar{y}_r) - \bar{Y} = \bar{Y}_r - \bar{Y} = QC . \tag{2.6}$$

The bias is determined by two factors: the relative size of the nonresponse stratum and the value of the contrast.

We call an estimator *precise* if its possible outcomes do not fluctuate very much. The precision can be measured by the variance of the estimator. If an estimator is biased, the precision is not a good indicator of the quality of the estimator. We call an estimator *accurate* if its possible outcomes are close to the value which is to be estimated. The accuracy is measured by the mean square error. It can be expressed as

$$\mathrm{MSE}(\bar{y}_r) = E(\bar{y}_r - \bar{Y})^2 = V(\bar{y}_r) + B^2(\bar{y}_r) , \tag{2.7}$$

in which $V(\bar{y}_r)$ denotes the variance of the estimator $\bar{y}_r$. In general the variance will decrease as the sample size increases. However, this is not necessarily the case for the bias component of the mean square error. As can be seen from (2.6), the bias-part of (2.7) is independent of the sample size. Increasing the sample size will not reduce the bias. Some other measure must be taken to decrease the bias.

3. Call-Back Criterion

Usually, the quality of an estimate of a population characteristic is indicated by a confidence interval. If all sampled elements cooperate and the sample mean $\bar{y}$ is approximately normally distributed, then the 95 %-confidence interval for $\bar{Y}$ is equal to I, where

$$I = (\bar{y} - 1.96\, S(\bar{y}),\ \bar{y} + 1.96\, S(\bar{y})) \tag{3.1}$$

in which $S(.)$ denotes the standard error of the estimator. The probability that this interval contains the true value $\bar{Y}$ is approximately equal to $P(\bar{Y} \in I)$, where

$$P(\bar{Y} \in I) = .95 . \tag{3.2}$$

In case of nonresponse, the construction of the confidence interval will have to be based on the mean $\bar{y}_r$ of the observations in the response stratum instead of $\bar{y}$. If $\bar{y}_r$ is substituted for $\bar{y}$ in (3.1), the resulting interval I_r has the property that

$$P(\bar{Y} \in I_r) = \phi(1.96 - B(\bar{y}_r)/S(\bar{y}_r)) - \phi(-1.96 - B(\bar{y}_r)/S(\bar{y}_r)) , \quad (3.3)$$

in which $\phi(.)$ is the standard normal distribution function. Table 2 gives values of the probability (3.3) for a number of values of $|B(\bar{y}_r)/S(\bar{y}_r)|$.

Table 2. The Confidence Level of the Confidence Interval I_r as a Function of $|B(\bar{y}_r)/S(\bar{y}_r)|$

$\lvert B(\bar{y}_r)/S(\bar{y}_r)\rvert$	$P(\bar{Y} \in I_r)$
.0	.95
.2	.95
.4	.93
.6	.91
.8	.87
1.0	.83
1.2	.78
1.4	.71
1.6	.64
1.8	.56
2.0	.48

It is clear that the confidence level can be much lower. If the bias is equal to the standard error the confidence level is .83 and not .95. As the ratio $B(\bar{y}_r)/S(\bar{y}_r)$ differs more from zero, the situation becomes worse and the confidence interval no longer gives a correct indication of the true value.

We apply these results in the construction of a call-back criterion. It is only a proposition. Other criteria are also possible. The underlying assumption of our criterion is that the bias is unduly large if

$$|B(\bar{y}_r)/S(\bar{y}_r)| \geqslant 1. \quad (3.4)$$

Since $V(\bar{y}_r)$ can be approximated by

$$(N/n - 1)\, \bar{Y}_r(1 - \bar{Y}_r)/N(1 - Q) , \quad (3.5)$$

and since (3.5) is smaller than or equal to

$$.25(N/n - 1)/N(1 - Q), \quad (3.6)$$

call-backs should take place if

$$|C| \geqslant .5\sqrt{(N/n)-1} \;/\; Q\sqrt{N(1-Q)} . \quad (3.7)$$

See also Kersten and Bethlehem (1981). The relative size Q of the nonresponse stratum can be estimated from the sample. Of course, no exact information for C is available. However, the inequality (3.7) can be useful because it allows an educated guess about C to be made on the basis of other nonresponse analyses or on experience with the specific survey. It is a useful tool in the planning and processing of a survey. Since it incorporates both sampling error and nonresponse bias, it helps to allocate financial resources in such a way that within budget constraints, the best possible estimates are obtained.

The call-back criterion was applied in the 1981 Labour Force Survey. From all municipalities a sample with the sampling fraction of .05 was selected. So, for each municipality n and N were known, and Q could be estimated by the nonresponse rate in the field. Using $N/n=20$ the call-back criterion becomes in this case:

$$|C| \geqslant .5\sqrt{19} \,/\, (Q\sqrt{N(1-Q)}\,). \quad (3.8)$$

Thus, call-backs should be made when (N, Q) lies above the curve, given by

$$\sqrt{N} = .5\sqrt{19} \,/\, (|C| Q \sqrt{1-Q}\,) , \quad (3.9)$$

for a given contrast C. On the basis of theoretical considerations and available knowledge about certain important and sensitive characteristics, realistic values of the contrasts were determined. The contrast-values were substituted in (3.9). Consequently, for each municipality that did not pass the criterion, call-backs were carried out.

Fig. 1 gives curves corresponding to a number of contrast values. A point on the curve corresponds to a nonresponse rate and the population size for which (3.9) holds. If the point, corresponding to a municipality, lies above the curve then call-backs must be made. In that case, the nonresponse bias is greater than or equal to the standard error of the estimate. Fig. 1 also contains the points for a number of large cities. If, for example, a contrast of 2 is considered, then call-backs must take place in Amsterdam when the nonresponse rate exceeds 15 percent and in Rotterdam when the nonresponse rate exceeds 20 percent. Higher nonresponse rates may be tolerated for smaller cities given the same considered contrast, since the standard errors of the estimates are larger there.

The criterion, as presented in (3.4), is meant to protect against improper interpretation of computed confidence intervals of estimates. In the case of small samples, the criterion allows a large bias. If a large sample is selected, only a very small bias is allowed. So, the criterion does not guarantee accurate estimates. If accuracy is to be taken into account, the right-hand part of Fig. 1 should not be used. Points in that part of the figure, which satisfy the criterion, correspond to inaccurate estimates.

4. The Basic Question Procedure

Hansen and Hurwitz (1946) were among the first to recognize that nonresponse can lead to biased estimates of population characteristics. To investigate the nonresponse in mail surveys, they proposed to select a sample of nonrespondents. The selected nonrespondents were visited by interviewers who tried to collect the desired information by means of a face-to-face interview. This technique can also be applied when the data in the original sample were collected by interviews. A sample of refusers were visited by interviewers specially trained to deal with nonrespondents.

A lot of information can be obtained by subsampling the nonrespondents. However, it is a costly and time-consuming technique, and therefore it is not often applied. Kersten and Bethlehem (1984) propose the Basic Question Procedure as an inexpensive alternative, which can be used if there is no hope of a complete interview (at that moment or later). The Basic Question Procedure is based on the idea that people who refuse to answer all questions, are frequently willing to answer one or two questions. Most surveys contain one or two basic questions that capture the essence of the survey as a whole. Table 3 presents the Basic Questions used for a number of Dutch Surveys. Answers of nonrespondents to Basic Questions give insight in possible bias with respect to important characteristics of the population.

The use of the Basic Question Procedure need not to be restricted to refusals. For not-at-homes and for other nonresponse categories, these questions can be answered afterwards by telephone or by mail. In these cases, attention should be paid to response effects.

To be able to compare answers on the Basic Question by respondents and nonrespondents, the basic question has to be asked to both groups in circumstances as similar as possible. For respondents, this means that the question should be one of the first in the interview. If the answer turns out to be wrong in the course of the interview, because it is inconsistent with answers to other questions, then the answer may not be corrected for reasons of comparison and bias reduction purposes. However, for other purposes the correct answer should also be processed.

If respondents and nonrespondents give different answers to the Basic Question, then this information can be used in two ways:

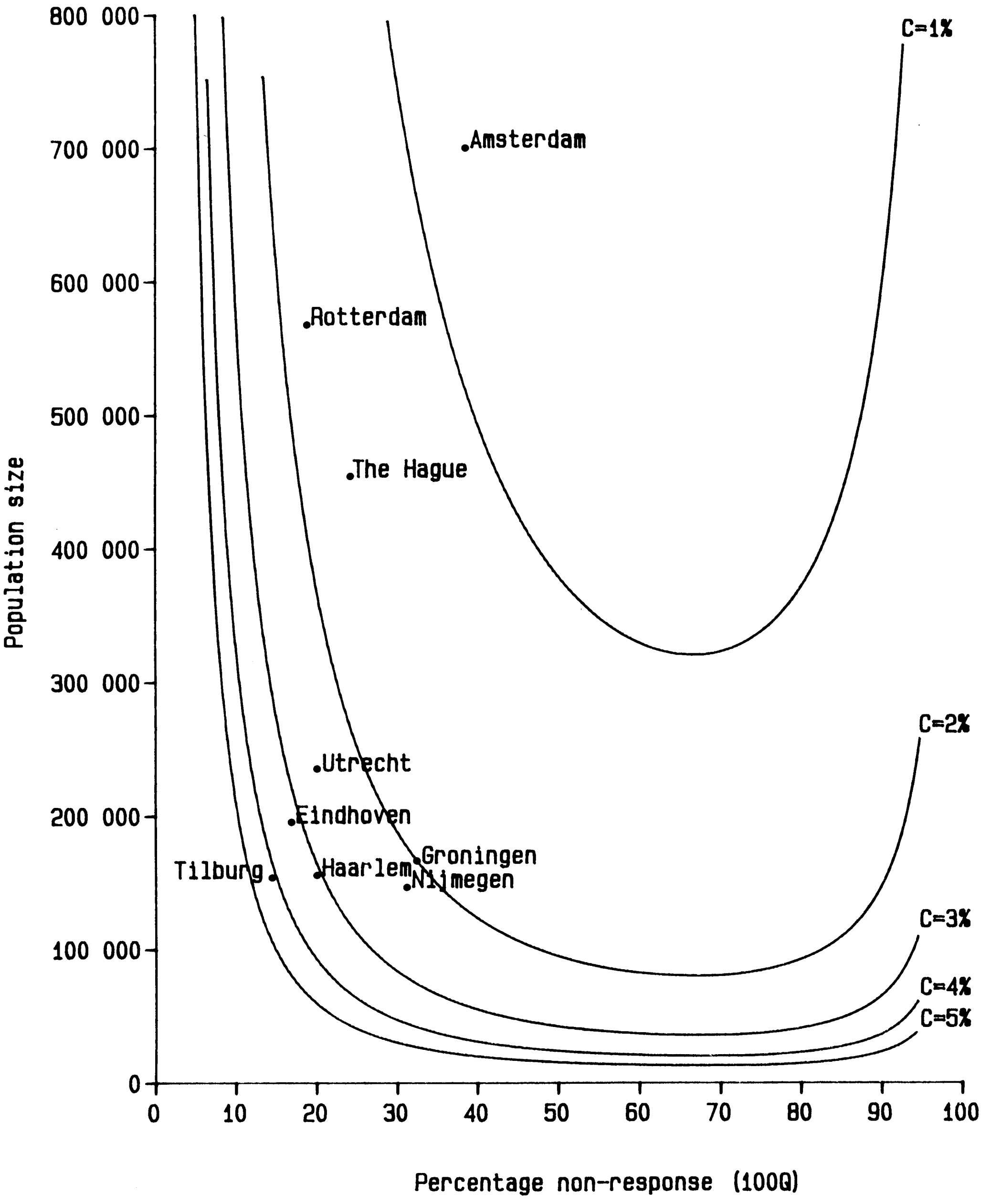

Fig. 1. Contrast Curves for a Number of Contrast Values and a Sample Size n=.05N

In the first place, the estimation procedure for the Basic Question itself can be improved. Suppose $\bar{x}_r$ is the mean score on the Basic Question of the n_r respondents and suppose $\bar{x}_{nr}$ is the mean for the n_{nr} nonrespondents. Without the Basic Question Procedure, we only have $\bar{x}_r$ to estimate the population mean $\bar{X}$. With the Basic Question Procedure a better estimator is defined by

Table 3. Basic Questions Used in CBS Sample Surveys

Sample survey	Basic question
1981 Housing Demand Survey	Do you intend to move within two years?
1981 and 1983 Labour Force Surveys	How many people in this household have a job?
1981 Holiday Survey	Have you been on holiday during the last 12 months?
1982 Fertility Survey	Taking into account your present circumstances, and your expectations of the future, how many children do you plan on having from this moment on?

$$\bar{x}_{BQ} = (n_r \bar{x}_r + n_{nr} \bar{x}_{nr})/n. \tag{4.1}$$

This estimator gives equal weights to the units, as is the case of simple random sampling. Under the assumption that nonrespondents answering the Basic Question are a random sample from all nonrespondents, the estimator $\bar{x}_{BQ}$ is an unbiased estimator of $\bar{X}$.

The Basic Question variable can also be used as an auxiliary variable in the estimation of other population characteristics. For example, if the Basic Question variable is continuous, then the regression estimator for the mean $\bar{Y}$ of some other characteristic becomes

$$\bar{y}_{BQ} = \bar{y}_r + b_r(\bar{x}_{BQ} - \bar{x}_r)\ , \tag{4.2}$$

where b_r is the estimate of the regression coefficient based on the response.

The Basic Question Procedure was applied in the 1981 Dutch Housing Demand Survey. Table 4 gives the answers of respondents and nonrespondents to the Basic Question. All respondents answered the Basic Question (the response rate after a maximum of three visits was 71.2%) whereas 35% of the nonrespondents answered the Basic Question. Obviously, respondents are more inclined to move than the nonrespondents that answer the Basic Question. Under the assumption that these nonrespondents do not differ significantly from the other nonrespondents, one may conclude that respondents are more inclined to move that nonrespondents.

After a few months a second phase survey was carried out in which a large group of nonrespondents was visited again (including nonrespondents who answered the Basic Question). For nonrespondents who answered all questions in the second phase (including the Basic Question) and who answered the Basic Question in the first phase we can compare their Basic Question answers. The results are presented in Table 5.

At first sight the answers are rather consistent. About 83% (=100x1355/1638) of the individuals stick to their answer. However, some critical remarks can be made. In the first phase 13.7% (=100x225/1638) of the individuals indicate that they intend to move. In the second phase response, this percentage is 20.8 (=100x340/1638). A possible explanation is that when a complete interview took place, for reasons of questionnaire design, the Basic Question was asked somewhere in the middle

Table 4. Answers to the Basic Question in the 1981 Dutch Housing Demand Survey

Do you intend to move within two years?	Respondents	Nonrespondents
Yes	17 606 (30%)	1 076 (13%)
No	41 674 (70%)	7 334 (87%)
Total	59 280 (100%)	8 410 (100%)

Table 5. Answers to the Basic Question in the First and Second Phase

Answer to the Basic Question in the first phase	Second phase answer yes	no	total	percentage
Yes	141	84	225	14
No	199	1 214	1 413	86
Total	340	1 298	1 638	100
Percentage	21	79	100	

of the interview after questions about present living conditions. The Basic Question is, in general, answered at the door in a rather difficult situation. It is sometimes not clear whether the person who gives the answer is the person who was supposed to be interviewed, whether he/she can give the exact information and whether the correct information is given. Furthermore, in the complete interview the respondent did not have the opportunity to answer "I don't know," whereas this answer was possible when answering the basic question only. Finally, there is a considerable time gap between the two phases. It is possible that people change their minds in the meantime, e.g., due to the survey they become more aware of their housing conditions. It is also possible that people have moved in the time interval between sample selection from the frame and the actual visit. For those who moved during the course of the survey, the second-phase visit took place at their new addresses. The inclusion of this group could have reduced the number of people who reported plans to move.

In this section only one category of nonrespondents was distinguished. As indicated by Kersten and Bethlehem (1984) the behaviour of different types of nonrespondents may be different. So, it might be worthwhile to divide the nonresponse stratum in a number of substrata (refusers, not-at-homes, etc.). The generalization of the theory is straightforward.

All in all, one may conclude that the Basic Question Procedure has some nice properties, but further research on the validity of the answers is necessary.

5. A General Weighting Technique

A frequently used technique to reduce the nonresponse bias is poststratification. Every observed object is assigned a weight, and estimators of population characteristics are obtained by processing the weighted observations instead of the observations themselves.

The idea behind poststratification (sometimes called stratification after selection of the sample) is to divide the population into homogeneous strata. If all elements within a stratum resemble each other, then stratum estimates are not very biased. Strata are constructed using auxiliary variables which preferably have a strong relationship with the characteristic to be investigated. All objects within a stratum are given the same weight such that the weighted sample distribution of the auxiliary variables agrees with the population distribution of these variables. If the relationships are strong enough, the weighted sample distribution of the characteristic to be investigated will resemble the population distribution.

Two problems may arise when weighting is carried out: strata without observations and lack of adequate population information about the auxiliary variables. To avoid these problems, Bethlehem and Keller (1983) developed a general weighting technique.

Suppose there are p auxiliary variables. Associated with each object k in the population are a value Y_k which is the target variable and a vector $\underline{X}_k = (X_k^{(1)}, X_k^{(2)},\ldots,X_k^{(p)})'$ of values of the auxiliary variables. If the auxiliary variables are correlated with the target variable, then for a suitably chosen vector $\underline{\beta} = (\beta_1, \beta_2,\ldots,\beta_p)'$ of regression coefficients the residuals

$$\varepsilon_k = Y_k - \sum_{h=1}^{p} \beta_h X_k^{(h)} = Y_k - \underline{X}_k'\underline{\beta} \tag{5.1}$$

will vary less than the values of the target variable itself. Application of ordinary least squares gives the optimal value

$$\underline{\beta} = \left[\sum_{k=1}^{N} \underline{X}_k\underline{X}_k'\right]^{-1} \left[\sum_{k=1}^{N} \underline{X}_k Y_k\right] \tag{5.2}$$

for $\underline{\beta}$. The estimator for $\underline{\beta}$ on the basis of a simple random sample is equal to

$$\underline{b} = \left[\sum_{i=1}^{n} \underline{x}_i\underline{x}_i'\right]^{-1} \left[\sum_{i=1}^{n} \underline{x}_i y_i\right], \tag{5.3}$$

in which $\underline{x}_1, \underline{x}_2,\ldots, \underline{x}_n$ are the vectors of auxiliary values belonging to the sampled objects. Using $\underline{b}$, the general regression estimator is defined as

$$\bar{y}_R = \underline{b}'\underline{\bar{X}}. \tag{5.4}$$

Bethlehem (1985) shows that this estimator is approximately unbiased if and only if there exists a vector $\underline{c}$ such that $\underline{c}'\underline{X}_k = 1$, for all k. In that case (5.4) can be rewritten as

$$\bar{y}_R = \bar{y} - \underline{b}'(\underline{\bar{x}} - \underline{\bar{X}}). \tag{5.5}$$

It can be shown that the use of the estimator implies that weights are assigned to sampled objects. The weight which is assigned to sample element i is given by

$$w_i = \underline{x}_i' \left[\sum_{i=1}^{n} \underline{x}_i\underline{x}_i'\right] \underline{\bar{X}} \quad . \tag{5.6}$$

Using techniques analogous to those of Cochran (1977) in his derivation of the variance of the simple regression estimator, the variance of the general regression estimator can be approximated by

$$V(\bar{y}_R) = \frac{1-f}{n}\frac{1}{N-1}\sum_{k=1}^{N}(Y_k - \underline{X}_k'\underline{\beta})^2$$

$$= \frac{1-f}{n}\frac{1}{N-1}\sum_{k=1}^{N}\varepsilon_k^2, \tag{5.7}$$

using the notation of (5.1), and in which $f=n/N$ is the sampling fraction. The sample based estimator of the variance is equal to

$$v(\bar{y}_R) = \frac{1-f}{n}\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \underline{x}_i'\underline{b})^2. \tag{5.8}$$

The general regression estimator is usually applied in the case of continuous auxiliary variables. However, it can just as well be applied if the auxiliary variables are qualitative. With qualitative variables, each auxiliary variable must be transformed into a set of dummy variables. If the foregoing theory is developed for a qualitative variable, the regression estimator can be used as an estimator for stratification after selection of the sample.

Suppose there is a qualitative auxiliary variable with p categories. To include this variable in a linear model, the variable is replaced by a vector of p dummy variables. Each category (stratum) has its corresponding dummy variable. It takes the value 1 if the object belongs to the particular stratum and 0 otherwise. So, for every element only one dummy takes the value 1; all other dummies take the value 0.

If simple random sampling is assumed, the estimator $\underline{b}$ for $\underline{\beta}$ turns out to be

$$\underline{b} = (\bar{y}^{(1)}, \bar{y}^{(2)},\ldots,\bar{y}^{(p)})', \tag{5.9}$$

in which $\bar{y}^{(h)}$ is the mean of the observations in stratum h, for $h=1,2,\ldots,p$. The vector of population means of the auxiliary variables is in this case equal to

$$\underline{\bar{X}} = (N_1, N_2, \ldots, N_p)'/N, \tag{5.10}$$

where N_h is the size of stratum h, for $h=1,2,\ldots,p$. By substitution of (5.9) and (5.10) in (5.4) the general regression estimator becomes

$$\bar{y}_{PS} = \sum_{h=1}^{p} N_h \bar{y}^{(h)}/N \,. \tag{5.11}$$

The subscript *PS* indicates that estimator (5.11) is equal to the ordinary poststratification estimator. In our general theory of weighting we call this estimator *oneway stratification.*

Poststratification can also be applied if several qualitative auxiliary variables are available. Suppose we have m such variables with the number of categories equal to $p_1, p_2, \ldots, p_m$. Every combination of values of the auxiliary variables forms a stratum, the total number of strata being equal to $p=p_1 \times p_2 \times \ldots p_m$. If the m qualitative auxiliary variables are replaced by p dummy variables then the situation is reduced to that of oneway stratification.

If the theory of linear models is restricted to use of qualitative independent variables, it is usually called analysis of variance. For this reason our terminology has its roots in the analysis of variance. The auxiliary variables correspond to factors and the strata to cells. Stratification in which strata are constructed on the basis of all possible combinations of values of the auxiliary variables corresponds to an analysis of variance in which the model contains the highest order interaction. Therefore we call this type of poststratification *complete multiway stratification*. If complete multiway stratification is applied two problems may occur:

The first problem is the problem of empty strata. If there is so much auxiliary information available that it allows complete multiway stratification, the resulting number of strata may be so large that there is a risk of empty strata in the sample. In daily practice this problem is usually solved by collapsing strata. As this is frequently done by hand, it is a time-consuming process.

The second problem is created by lack of auxiliary information. Sometimes auxiliary information is available, but is not detailed enough to allow complete multiway stratification. For instance, complete multiway stratification by sex, age, marital status and region requires knowledge of the population totals for each combination of sex, age, marital status and region. If the totals are known for each combination of sex, age, marital status and separate totals for each region, then complete stratification is not possible.

Incomplete multiway stratification offers a solution for the problems described above. If the highest order interactions are removed from the model and replaced by lower order interactions, then, in many cases, the problems disappear.

When describing an incomplete multiway stratification, it is convenient to use a simple notational language. Complete multiway stratification by sex, age, marital status and region is denoted by SEX x AGE x MARITAL STATUS x REGION. Incomplete multiway stratification which uses the population totals for each combination of sex, age and marital status on the one hand and the population totals for each region on the other hand is denoted by (SEX x AGE x MARITAL STATUS) + REGION. One set of dummy variables indicates the combinations of sex, age and marital status, and another set of dummy variables denotes the region.

Stratification is reduced to estimation of the parameter vector $\underline{\beta}$. The number of parameters to be estimated is smaller in incomplete stratification than in complete stratification. So, incomplete stratification decreases estimation problems. For instance, if sex has two categories, age ten categories, marital status four categories and region eleven categories, then (SEX x AGE x MARITAL STATUS) x REGION comes down to estimation of 880 parameters, whereas (SEX x AGE x MARITAL STATUS) + REGION requires at most 91 estimated parameters. On the other hand, the model behind the incomplete stratification might not fit as well as the model behind the complete stratification. However, we believe that in practical situations the incomplete stratification model is based on enough parameters to give as good, or nearly as good, a fit as the complete stratification model.

We illustrate our approach by an example based on data from the 1977/1978 Dutch Housing Demand Survey. Our aim is to estimate the mean household income. Persons with item nonresponse on the income question are treated as nonrespondents. For weighting purposes three auxiliary variables are available: age (6 categories), sex (2 categories) and marital status (2 categories). A complete multiway stratification by age, sex and marital status (denoted by AxSxM) leads to 6x2x2=24 strata. When the sample size in some strata is too small or when population totals are missing for each combination of age, sex and marital status, an alternative weighting scheme may be considered. To be able to use as much auxiliary information as possible a weighting scheme can be specified in such a way that the weighted marginal distributions of sex, age and marital status agree with the marginal population distributions. So, three sample distributions are simultaneously fitted to their population equivalents. We denote this weighting scheme by A+S+M.

Schemes AxSxM and A+S+M are only two possible weighting schemes, using age, sex and marital status. Many more are possible. Table 6 lists all possible weighting schemes using age, sex and marital status (or subsets of them), and the corresponding estimates of the mean income and its (estimated) standard error. For example (SxA)+(AxM) means that the weighted sample totals of sex by age are equal to the corresponding population totals and that the weighted sample totals of age by marital status are equal to their population equivalents.

A striking phenomenon is the shift of the estimate of the mean income. The estimate increases as more auxiliary information is used. Without weighting, the value of the estimate is 23 494 and with weighting the estimates can be 24 000 or more. Further, it can be observed in the table that the precision of the estimator increases as more auxiliary information is used. The standard error decreases from 182 (no weighting) to 152 (extensive weighting). These phenomena can be seen as indications that weighting increases precision and decreases nonresponse bias. More details about this example and the general weighting technique can be found in Bethlehem and Keller (1983).

The theory developed in this section assumes complete response. With these assumptions, it is clear that proper use of auxiliary information improves the accuracy of estimates. The theory can be extended to incorporate nonresponse. Bethlehem (1985) develops the theory of the general regression estimator under the Random Response Model. In his paper, he shows that estimators that make use of well-chosen auxiliary variables, i.e., auxiliary variables that account for the behaviour of the target variable of the survey for a large extent, have a smaller bias than estimators that do not use auxiliary information. Hence, in the case of weighting we can draw two important conclusions; proper weighting improves accuracy and reduces the bias.

Table 6. Estimates of the Mean Household Income in the 1977/1978 Dutch Housing Demand Survey

Weighting Scheme[1]	Estimate	Standard Error
None	23 494	182
S	23 613	179
A	23 990	170
M	23 624	161
SxA	24 012	167
S+A	24 065	168
SxM	23 809	160
S+M	23 675	160
AxM	23 987	153
A+M	24 071	154
S+A+M	24 104	154
(SxM)+A	24 172	153
(SxA)+M	24 078	153
(AxM)+S	24 004	152
(SxA)+(SxM)	24 149	153
(AxM)+(SxM)	24 076	152
(SxA)+(AxM)	23 985	152
(SxA)+(AxM)+(SxM)	24 054	152
SxAxM	24 048	152

[1] S=sex, A=age, M=marital status.

6. Conclusions

In this paper we have discussed three techniques to deal with nonresponse: a call-back criterion, the Basic Question Procedure and a general weighting technique.

Sometimes it is said that every sample survey with a nonresponse rate exceeding 15% should be thrown into the waste-basket. In the construction of the call-back criterion, we advocate a more subtle point of view. A high nonresponse rate should not be fatal if it is still possible to make valid confidence intervals. This naturally leads to a criterion which incorporates both standard error and bias. As mentioned before, our criterion does not necessarily produce accurate estimates. This should be based on other criteria or an extended version of our call-back criterion.

The Basic Question Procedure is a quick and inexpensive method to collect information on nonrespondents. Although intuitively attractive and easy to implement, there are some problems which should be solved. If the same question is asked to a cooperative respondent and a potential nonrespondent, one may wonder if some kind of measurement error is not introduced. A second problem concerns the representativeness of the nonrespondents who answer the Basic Question. Are they typical respondents, typical nonrespondents or something in between? Further research to prove the usefulness of the procedure is necessary.

The general weighting technique is a useful extension of ordinary weighting techniques. On the one hand, it avoids some of the problems encountered in ordinary weighting, and on the other hand, it uses as much auxiliary information as possible; precision is improved and bias reduced. The theory is implemented in a general purpose computer program. In a number of Dutch sample surveys the technique has proved to be useful.

7. References

Bethlehem, J. G. (1985): The Non-Response Bias of Some Estimators. Internal CBS-report, Netherlands Central Bureau of Statistics, Voorburg.

Bethlehem, J. G. and Kersten, H. M. P. (1981a): Graphical Methods in Non-Response Analysis and Sample Estimation. Statistical Studies 31, Staatsuitgeverij, The Hague.

Bethlehem, J. G. and Kersten, H. M. P. (1981b): Non-Respons en Post-Stratificatie. Kwantitatieve Methoden 4, pp. 21–44.

Bethlehem, J. G. and Keller, W. J. (1983): A Generalized Weighting Procedure Based on Linear Models. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 70–75.

Cochran, W. G. (1977): Sampling Techniques. Third Edition, Wiley, New York.

Hansen, M. H. and Hurwitz, W. H. (1946): The Problem of Nonresponse in Sample Surveys. Journal of the American Statistical Association 41, pp. 517–529.

Kalsbeek, W. D. (1980): A Conceptual Review of Survey Error due to Nonresponse. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 131–136.

Kersten, H. M. P. and Bethlehem, J. G. (1981): Het vaste responsmodel. Internal CBS-report, Netherlands Central Bureau of Statistics, Voorburg.

Kersten, H. M. P. and Bethlehem, J. G. (1984): Exploring and Reducing the Non-response Bias by Asking the Basic Question. Statistical Journal of the UN Economic Commission of Europe 2, pp. 369–380.

Lindström, H., Wretman, J., Forsman, G. and Cassel, C. (1979): Standard Methods for Non-Response Treatment in Statistical Estimation. Memo, Statistics Sweden.

Received January 1985
Revised June 1985