Journal of Official Statistics
Vol 4. No. 4, 1988, pp. 363–374

Comparison of Measurement Errors for Telephone Interviewing and Home Visits by Misclassification Models

Gad Nathan and Tami Eliav[1]

Abstract: The assessment of measurement errors due to the mode of data collection (home visit or telephone interview) can be carried out on the basis of repeated observations, for instance via panel surveys, for characteristics which do not change over time. A misclassification model is proposed for this case with probabilities of correct classification and of misclassification depending on the mode of collection. The estimation and comparison of these parameters is used to assess the effects of mode of collection.

The models proposed are applied to empirical data from the Israel Labour Force Survey, in which home visits and telephone interviews are used in different rounds of the survey for the same units. The effects of mode of collection on model parameters are assessed for educational characteristics which are not subject to change over time (for those who no longer study). This is done separately for the subpopulation of households who have telephones and are willing to respond by telephone and for the subpopulation of households who respond only by home visit. The results indicate that responses obtained by telephone interview have higher consistency than those obtained by home visits.

Key words: Misclassification; response errors; mode of collection.

1. Introduction

The rising costs of survey field-work, especially when carried out by face to face home interviews, have led many statistical agencies to consider alternative methods of data collection. Sample selection via random digit dialing (RDD) (Waksberg (1978)) and the use of computer-assisted telephone interviewing (CATI) (Nicholls and Groves (1985)) have greatly increased the benefits of telephone surveys. There is no doubt that sampling and interviewing by telephone is fast becoming "a major development in the history of survey methods" (Groves and Kahn (1979)).

In most countries outside North America, the use of telephone interviewing has developed more slowly. In the United Kingdom Collins (1983) found that 56 % of market research agencies used telephone interviewing and Christoffersen (1984) reports increasing use of telephone interviewing in the Scandi-

[1] Department of Statistics, Hebrew University, 91905 Jerusalem, Israel.

Acknowledgements: We wish to thank Malka Kantorowitz, Director of the Statistical Methods Division, Central Bureau of Statistics, for her help with useful comments and ideas on an earlier version of this paper presented at the ISI session in Tokyo (Nathan (1987)). Thanks are also due to David Sageev, Deputy Director of the Division, for his help in preparation of the data.

navian countries. However, the use of the telephone for surveys outside of North America is still at the most partial or supplementary. Its primary use in many cases is for follow-ups, contacting not-at-homes and for screening. The telephone is also widely used in surveys based on non-probability samples (e.g., quota sampling).

There are several reasons for the limited use of telephone interviewing outside North America. The main reason is that telephone coverage of households is far from complete in most countries, especially for the rural population. In the United Kingdom an estimated 73 % of households had telephones (Sykes and Hoinville (1985)). In Israel only 69 % of households had telephones in 1983 – 70 % in urban localities and only 53 % in rural areas (Central Bureau of Statistics (1985)). Thus any probability sample survey would of necessity be a mixed-mode survey, with telephone interviewing supplemented by home visits or mail follow-up for the non-telephone households. Furthermore, in many countries, especially in Europe, the use of the telephone is often still limited to important business and social matters and its use for interviewing is perceived as an unwarranted intrusion of privacy. Another reason is the feeling, often expressed by field staff, that response to telephone interviews is in some way less accurate than that obtained by face to face interview in the respondent's home. There seems to be little empirical evidence to support this claim and several studies have found only small differences between responses obtained by different modes of collection (Colombotos (1969); Kantorowitz and Nathan (1987); Klecka and Tuchfarber (1978); Rogers (1976); Schuman and Presser (1981)).

Simple comparisons between responses obtained by telephone surveys with those obtained from home visits are difficult to assess because of confounding between the effects of mode of collection (telephone interview or home visit) and the effect due to differences in characteristics between households without telephones (and those not willing to answer by telephone) and households with telephones who allow telephone interviewing. Well designed experiments of the split ballot type, in which subjects are randomly allocated to mode of collection, partially overcome this difficulty, but are difficult to implement with effective control (Hochstim (1967); Locander, Sudman, and Bradburn (1974); Rogers (1976); and Wiseman (1972)). Moreover, while it is relatively simple to limit the comparison to households with telephones, it is more difficult to neutralize the effect of unwillingness to be interviewed by telephone. The answers to a hypothetical question on such willingness in a home interview must be treated with extreme caution.

Obviously, "pure" response error effects of mode of collection can best be assessed by independent measurements for the same subject by telephone and by home interview. This has been done by telephone reinterviews of subjects who have previously responded to a home interview (e.g., Rogers (1976)). Simple reinterviewing will allow a comparison but will not provide an assessment of the measurement errors for which, in general, at least two measurements for each mode are required for each subject. There is an inherent difficulty in attaining independent observations from the same subject over a short period of time. Here we consider the possibility of using a multi-round survey, with relatively long time intervals between interviews to assess the effect of mode of collection on characteristics which are invariant over time, at least over the period of the survey (such as last school attended for those not studying during the survey period). This is an extension of the general treatment of response error micro-effects from repeated surveys, given in Kantorowitz and Nathan (1987), without a mode of collection effect.

In Section 2 a misclassification model is proposed, from which, under certain assumptions, both "true" class proportions and misclassification probabilities can be estimated, both for responses obtained by telephone interview and for responses obtained by home interview for the same set of respondents. These misclassification probabilities can be assessed for the set of respondents who exhibit willingness to respond via telephone interviews (by actually participating in at least one telephone interview), and can be compared with those estimated for the set of respondents from whom only home interviews are obtained (due to the lack of a telephone or to their unwillingness to be interviewed over the telephone).

In Section 3 the model is applied to data from the multi-round Israel Labour Force Survey for educational characteristics, which are invariant over time for the subpopulation of those not studying during the survey period. The survey uses both telephone interviews and home visits for each subject, if possible. Thus, the effects of mode of collection and of lack of telephone (or unwillingness to be interviewed by telephone) on the biases in the distributions can be investigated.

2. The Misclassification Model

We extend the misclassification model proposed in Kantorowitz and Nathan (1987) in order to deal with the effects of mode of collection for qualitative variables. Previous work on misclassification, under different assumptions, was done by Press (1968), Hochberg (1977), and Cowan (1985).

We consider the population as being divided by a qualitative polytomous variable into categories with an unknown probability, R_k, of belonging to category k, $(k=1,\ldots,c$ and $\sum_k R_k = 1)$. At each round of a multi-round survey, each unit is classified (correctly or not) into one of the c categories.

Let $P^{(j)}_{kk'}$ denote the conditional probability that a unit reports its category as k' by mode of collection $j(=1,\ldots,m)$, given that its true category is $k(k,\ k'=1,\ \ldots,\ c$ and $\sum_{k'} P^{(j)}_{kk'} = 1)$. We assume that the probabilities, $P^{(j)}_{kk'}$, of misclassification to category k' (for $k' \neq k$) and of correct classification (for $k' = k$) are constant over rounds for the same mode of collection, j. We also assume that classification is independent (conditionally, given the true category) over rounds.

Let x_{jk} be the number of times a given unit is classified as belonging to category k by mode of collection j, $(j=1,\ \ldots,\ m;\ k=1,\ \ldots,\ c)$; let $\mathbf{x}=(x_{11},\ \ldots,\ x_{1c},\ \ldots,\ x_{m1},\ \ldots,\ x_{mc})$ be the unit's mc-component observation vector; and let $r_j(\mathbf{x}) = \sum_k x_{jk}$ be the number of rounds the unit is observed by mode of collection j. Then, under the above assumptions, the probability of observing $\mathbf{x}$ can be expressed as:

$$\pi(\mathbf{x}) = \textstyle\sum_k R_k \prod_j r_j(\mathbf{x})!\ \prod_{k'} \frac{1}{x_{jk}!}\,[P^{(j)}_{kk'}]^{x_{jk'}}. \qquad (1)$$

If we assume that a simple random sample of n independent observations on $\mathbf{x}$ is obtained, then the distribution of the frequencies of observing $\mathbf{x}$, $f(\mathbf{x})$, is multinomial with probabilities $\pi(\mathbf{x})$, where $\sum_{\mathbf{x}} f(\mathbf{x}) = n$ and $\sum_{\mathbf{x}} \pi(\mathbf{x}) = 1$. Thus the kernel of the log likelihood of the observations is given by:

$$\ln L = \textstyle\sum_{\mathbf{x}} f(\mathbf{x}) \ln \pi(\mathbf{x}). \qquad (2)$$

The maximum likelihood estimators of the parameters R_k and $P^{(j)}_{kk'}$ must be solutions of the following equations, obtained by equating the appropriate partial derivatives of the Lagrange function for (2) to zero:

$$R_k = n^{-1} \sum_{\mathbf{x}} f_k(\mathbf{x})$$
$$P^{(j)}_{kk'} = \frac{\sum_{\mathbf{x}} x_{jk'} f_k(\mathbf{x})}{\sum_{\mathbf{x}} r_j(\mathbf{x}) f_k(\mathbf{x})} \tag{3}$$

where

$$\pi_k(\mathbf{x}) = R_k \Pi_j r_j(\mathbf{x})!\ \Pi_{k'} \frac{1}{x_{jk}!} \left[P^{(j)}_{kk'}\right]^{x_{jk'}}. \tag{4}$$

is the joint probability of observing $\mathbf{x}$ and of the true category being k and

$$f_k(\mathbf{x}) = f(\mathbf{x})\, \pi_k(\mathbf{x}) / \pi(\mathbf{x}) \tag{5}$$

is the expected frequency of observations $\mathbf{x}$ for which the true category is k, given that the observed total frequency of $\mathbf{x}$ is $f(\mathbf{x})$.

To ensure unique solutions of (3), which are then MLE, the number of independent observations, $\mathbf{x}$, must not be less than the number of independent unknown parameters, which is $(mc+1)\ (c-1)$ with $c-1$ independent values of R_k and $mc(c-1)$ independent values of $P^{(j)}_{kk'}$. We shall also require, for identifiability, that

$$P_{kk} \geq P_{kk'} \text{ for } k \neq k'. \tag{6}$$

Explicit solutions of (3) are, in general, not available but a simple iterative procedure is obtained by substituting the values of $f_k(\mathbf{x})$, computed by (3) and (4) from one set of parameter values, into (5) to obtain the next set. For initial values of the parameters the following "naive" estimators can be used

$$\hat{R}_k = n^{-1} \sum_{\mathbf{x}} f(\mathbf{x}) \frac{\sum_j x_{jk}}{\sum_j r_j(\mathbf{x})},$$
$$\hat{P}^{(j)}_{kk'} = n^{-1} \hat{R}_k^{-1} \sum_{\mathbf{x}} \frac{f(\mathbf{x}) x_{jk} x_{jk'}}{r_j(\mathbf{x})[r_j(\mathbf{x}) - 1]} \tag{7}$$

for $k' \neq k$, and

$$\hat{P}^{(j)}_{kk} = 1 - \sum_{k' \neq k} {}_{k'}\, \hat{P}^{(j)}_{kk'}.$$

This procedure can be shown to be equivalent to the EM algorithm (Dempster, Laird, and Rubin (1977)). The condition $r_j(\mathbf{x}) \geq 2$ is required to ensure identifiability and thereby unique solutions of (3), subject to (6), which are MLE.

3. The Empirical Results

The misclassification model described above was applied to a set of data on educational characteristics derived from four rounds of a single panel of the Israel Labour Force Survey. The data set was derived for the general analysis of response error effects in Kantorowitz and Nathan (1987) which describes the data set's characteristics and sources in detail. A brief summary of the aspects of the data set most germane to the present analysis follows.

As pointed out previously, the models can be applied only to characteristics which are invariant over time. Most of the variables measured in Labour Force Surveys or other current surveys either change over time (e.g., labour force characteristics), or are obtained only at the first interview (e.g., date of birth). In the Israel Labour Force Surveys the following set of educational questions is asked independently in each of the rounds (up to four) for each household and each individual aged 14 and over.

For those who did not attend school during the whole period of the survey (ascertainable via question 8), the variables "years of study" (question 9) and "type of last school attended" (question 10) can be regarded as invariant over the rounds in which the respondent participated.

The Israel Labour Force Survey is a current rotating panel survey with four panels investigated each quarter for urban localities

8. Have you attended school or do you attend school now?

   1 Attended only in the past
   2 Presently attending (even if on vacation)
   3 Never attended school -----> skip to question 11

9. For how many years have you attended school? □□

10. What is the type of school last attended?

   01 Primary school
   02 Intermediate school
   03 Vocational or agricultural secondary school
   04 Secondary school
   05 Yeshiva
   06 Teacher training college
   07 Technical post-secondary school
   08 Other post-secondary school
   09 Academic institution
   10 Other, specify

(Central Bureau of Statistics (1987)). Each dwelling unit in the survey is investigated for two consecutive quarters and after a break of two quarters, for two additional consecutive quarters. The sample design is stratified single-stage stratification for large localities with each locality serving as a stratum, and two-stage for smaller localities, with stratified PPS selection of localities in the first stage. The selection of dwelling units within selected localities is random systematic from lists of units, with equal final inclusion probabilities. The final sample for urban localities can be considered an approximate simple random sample of dwelling units. Since all persons aged 14 and over in selected dwelling units are investigated, the sample of individuals is clustered.

Face to face home interviews are carried out for most households in the first and last rounds. Telephone interviews are often used in the second and third rounds, at the discretion of the interviewer and after receiving the household's consent during the first interview. A small number of responses are obtained by mail. The respondent rule allows any adult member of the household to answer for all others. Overall non-response runs to about 13 %. The data set was based on the population in urban Jewish localities (about 86 % of the total) who participated in the panel first investigated in the last quarter of 1980. Only the 4 084 persons matched by survey identity number, sex and age and who reported consistently that they attended school only in the past (question 8) were retained.

The distribution of responses obtained from these persons in each round, by mode of collection, is given in Table 1.

Table 1. Responses by mode of collection in each round

Round	Total	Mode of collection		
		Home visit	Telephone	Mail
1	3 105	3 020	67	18
2	3 216	1 767	1 435	14
3	3 109	1 554	1 542	13
4	3 065	2 932	110	23

For the present analysis, only respondents who participated in at least two rounds were included, to ensure identifiability, and mail responses were excluded. In addition, 69 persons who responded only by telephone interview were excluded. The final data set included 11 901 responses for 3 435 persons.

This sample was divided into two: (1) those who responded at least once by home visit and at least once by telephone interview – "telephone households" – representing the population of households with telephones who could be reached and were willing to respond both by home visit and by telephone; and (2) those only responding by home visits – "non-telephone households" – representing the population of households without telephone or unwilling to respond by telephone. The breakdown is given in Table 2.

It should be emphasized that responses obtained over the different rounds for the same unit (whether by home visit or by telephone) were usually obtained by the same interviewer. This implies that the confounding between interviewer effect and that of mode of collection, or round is limited. This fact could, on the other hand, cause between-round response dependence. However, it should be noted that the time lags between rounds, ranging from three to fifteen months, are such that this dependence must be very small. Clustering of units (an average of 2.7 persons per household) may cause some departure from the assumption of independence between the units.

The results of applying the model of Section 2 to each of the subpopulations are given in Table 3 for groups of years of study and in Table 4 for last school attended. The results for groups of years of study shown clearly that, for telephone households, telephone interviewing results in less misclassification than home visits. For all four categories the probabilities of correct classification (underlined) are higher for telephone interview than for home visit, the greatest difference being that for the group with 9–10 years of study (more than 9 %). The overall probability of misclassification, $\sum_k R_k\left[1 - P^{(j)}_{kk}\right]$, for the telephone households is

Table 2. Persons responding (by home visit or telephone) in two or more rounds and their responses by mode of collection[1]

	Persons	Responses		
		Total	Home visit	Telephone
Total	3 381	11 901	8 896	3 005
"Telephone households" (both modes)	1 840	6 696	3 691	3 005
"Non-telephone households" (only home visits)	1 541	5 205	5 205	.

[1] Responses by mail and those responding only by telephone excluded.

Table 3. Estimates of misclassification probabilities for groups of years of study by type of household and by mode of collection (percentages)

True category k	Estimated percentage $\hat{R}_k$	Estimated conditional probability of reporting category k' – $\hat{P}^{(j)}_{kk'}$			
		1–8	9–10	11–12	13+
Telephone households – response by telephone interview					
1–8	28.2	94.7	3.6	1.3	0.3
9–10	16.2	5.1	85.2	8.6	1.1
11–12	32.5	1.5	1.1	94.4	2.9
13+	23.1	0.2	0.4	1.8	97.6
Telephone households – response by home visit					
1–8	28.2	90.5	5.3	3.4	0.7
9–10	16.2	10.1	75.9	12.6	1.5
11–12	32.5	1.5	4.4	89.5	4.6
13+	23.1	0.0	0.9	4.1	95.0
Non-telephone households – (response by home visit)					
1–8	30.8	95.5	3.0	1.4	0.2
9–10	21.7	9.2	80.9	8.8	1.2
11–12	29.7	2.8	5.2	87.8	4.2
13+	17.9	0.9	0.8	4.8	93.5

Table 4. Estimates of misclassification probabilities for last school attended by type of household and by mode of collection (percentages)

True category k	Estimated percentage $\hat{R}_k$	Estimated conditional probability of reporting category k' – $\hat{P}^{(j)}_{kk'}$			
		Primary	Vocational	Secondary[1]	Academic
Telephone households – response by telephone interview					
Primary	28.8	92.9	1.7	5.3	0.1
Vocational	23.7	3.0	86.7	9.9	0.4
Secondary[1]	30.7	2.7	5.5	89.4	2.4
Academic	16.8	0.2	0.4	5.0	94.4
Telephone households – response by home visit					
Primary	28.8	90.4	3.5	5.9	0.2
Vocational	23.7	5.2	80.8	13.8	0.1
Secondary[1]	30.7	3.2	7.1	87.9	1.7
Academic	16.8	0.2	0.4	6.6	92.9
Non-telephone households – (response by home visit)					
Primary	33.0	93.9	2.2	4.0	0.0
Vocational	28.2	4.7	84.3	10.6	0.4
Secondary[1]	28.8	4.3	8.9	84.8	2.1
Academic	10.0	0.0	1.0	2.7	96.3

[1] Includes post-secondary schools.

estimated as 11.1 % for home visits as against 6.3 % for telephone interviews. For non-telephone households the overall probability of misclassification (for home visit) is 10.2 %, somewhat lower than that for telephone households, but still much higher than for telephone interviews (for telephone households).

Table 5. Asymptotic confidence intervals for overall misclassification probabilities (telephone households – percentages)

	Years of study	Last school attended
Home visit	11.1 ± 4.7	12.2 ± 4.7
Telephone interview	6.3 ± 5.0	9.6 ± 4.8
Difference	5.8 ± 4.5	2.6 ± 3.9

Similar but less striking results are obtained for last school attended, but overall the advantage of telephone interviewing is still clear, with its estimated misclassification rate of 9.6 %, against 12.2 % for home visits. For non-telephone households the overall misclassification is again intermediate – 11.2 %. Correct classification probabilities range from a low of 84.3 % for vocational school to a high of 96.3 % for academic (both higher than those for home visits to telephone households).

The estimates above are, of course, subject to sampling errors. Estimates of the asymptotic variances of the sample estimates and their covariances were obtained, under the assumption of simple random sampling, from the estimated information matrix. While the separate estimates of the conditional misclassification probabilities and of the conditional probabilities of correct classification suffered from high sampling errors (up to 15 % relative standard error), sampling errors of the overall misclassification probabilities were small. Thus 95 % asymptotic confidence intervals for the overall misclassification probabilities are shown in Table 5. They indicate a significant difference between modes for "years of study," but not for "last school attended."

In addition, the difference between the two modes of collection was tested by the likelihood ratio test – the null hypothesis being that the probabilities $P^{(j)}_{kk'}$ do not depend on j. This hypothesis was clearly rejected at significance levels of less than 0.001, both for "years of study" and for "last school attended." The values of $-2ln(\lambda)$ (where λ is the likelihood ratio) were 6 397 and 2 787, respectively – with 12 degrees of freedom.

The misclassification model provides, besides estimates of misclassification probabilites, also estimates of the true proportion of units in each category, $\hat{R}_k$. These are given in the first column of Tables 6 and 7, separately for telephone households and for non-telephone households.

As might be expected, telephone households have a somewhat higher educational level than non-telephone households. Thus telephone households have higher proportions than non-telephone households in the categories "11+ years of study" and for "last school academic and secondary."

The distribution estimated by taking misclassification into account can be compared with the expectation of the distribution as reported by each mode of collection and for each subpopulation, $R^*_k(j) = \sum_{k'} \hat{R}_{k'} \hat{P}_{k'k}$, and the relative biases of the estimates of proportions can be estimated by: $[R^*_k(j)/\hat{R}_k] - 1$.

The results are shown in Tables 6 and 7 and they show, for the subpopulation of telephone households, the superiority of telephone interviewing over home visits, with respect to biases in estimating a distribution, under the assumptions of the model. Absolute relative biases in the proportions which would have been obtained by telephone are lower than those which would have been obtained by home visits in all but one category of "years

Table 6. Effect of mode of collection on biases in distribution by groups of years of study

Category	Estimated percentage	Expected percentage reported by		Relative bias (percentage)	
		Telephone interview	Home visit	Telephone interview	Home visit
Telephone households					
1–8	28.2	28.1	27.6	−0.5	−2.0
9–10	16.2	15.3	15.4	−5.8	−4.8
11–12	32.5	32.0	33.0	+1.1	+1.6
13+	23.1	23.8	23.9	+2.8	+3.4
				(2.6)[1]	(2.9)[1]
Non-telephone households					
1–8	30.8	·	32.4	·	+5.2
9–10	21.7	·	20.2	·	−7.1
11–12	29.7	·	29.3	·	−1.4
13+	17.9	·	18.3	·	+2.3
					(4.0)[1]

[1] Average, absolute values.

Table 7. Effect of mode of collection on biases in distribution by last school attended

Category	Estimated percentage	Expected percentage reported by		Relative bias (percentage)	
		Telephone interview	Home visit	Telephone interview	Home visit
		Telephone households			
Primary	28.8	28.3	28.3	−1.6	−1.8
Vocational	23.7	22.8	22.4	−3.8	−5.5
Secondary	30.7	32.2	33.1	+4.8	+7.7
Academic	16.8	16.7	16.2	−0.5	−3.5
				(2.7)[1]	(4.6)[1]
		Non-telephone households			
Primary	33.0	.	33.6	.	+1.7
Vocational	28.2	.	27.2	.	−3.7
Secondary	28.8	.	29.0	.	+0.7
Academic	10.0	.	10.3	.	+3.5
					(2.4)[1]

[1] Average, absolute values.

of study," and in all categories for "last school attended." Overall there is a reduction in the mean absolute relative bias for both variables – from 2.9 % to 2.6 % for "years of study" and from 4.6 % to 2.7 % for "last school attended."

Since telephone households represent only about a half of all households, an attempt was made to assess the overall impact of using telephone interviews for telephone households. The expected proportions were estimated separately for non-telephone households and the resultant biases estimated as above. The results are shown in the lower parts of Tables 6 and 7 and show that the biases for non-telephone households are neither consistently higher nor lower than those attained by home visits from telephone households.

The estimated proportions, $\hat{R}_k$, and the expected proportions $R^*_k(j)$ for the subpopulations were combined to obtain estimates for the total population, by weighting according to the sample representation (54 % – telephone households). The results still indicate the superiority of telephone interviews even if only used partially (i.e., a combination of telephone interviews for telephone households and of home visits for non-telephone households). Thus for "last school attended" absolute biases for telephone interviewing are still consistently lower than for home visits – with a mean absolute bias of 1.8 % against 2.6 %.

4. Discussion

Some of the obvious limitations of the simple misclassification model require further attention. One of the basic assumptions is the constancy of the category probabilities over the rounds. Since the characteristics investigated in our example are, by definition, invariant over time, the true category probabilities are certainly constant. However, there might be a differential bias in response that is not covered by our model. Thus several studies of the U.S. Current Population Survey (CPS) have indicated the importance of the phenomenon known as "rotation group bias" (see e.g., Bailar (1975)) which manifests itself in systematic differences in reports on the same characteristic over different rounds. A detailed analysis of our data on the basis of a linear model (Kantorowitz and Nathan (1987)) showed no significant differences between rounds for the educational characteristics, even when mode of collection was taken into account. Although a comparison of aggregates over a longer period did indicate some round effects, these could be attributed to differential non-response.

The effect of the assumption of independence of response deviations over rounds is more difficult to assess. Reinterview studies of the U.S. Current Population Survey have shown a small degree of dependence (a correlation coefficient of response deviations of about 0.3) for labour force characteristics (O'Muircheartaigh (1986)). However, practically all the CPS reinterviewing is carried out within two weeks of the original interview, whereas in our case the time between interview and reinterview (by the same mode) was mostly from 9 to 15 months and in all cases not less than 3 months. Furthermore, lower correlations could be expected for educational characteristics than for labour force characteristics. It is difficult to believe that substantial correlations could result under our conditions, but even if they did, they would effect the conclusions on the difference in quality between telephone and home interviewing only if there is a substantially higher correlation for telephone interviewing than for home visits. Although a model with correlated response deviations could be set up (see, e.g., Cowan (1985)) the number of parameters would be too large for viable estimation.

The Israel Labour Force Survey allows any adult respondent to answer for all household members and more than half the responses are indeed by proxy. The results of Kantorowitz and Nathan (1987) do indicate higher consistency by self-response than for proxy response (for heads of households). Thus, it is possible that the higher rate of self-response in telephone interviewing than in home visits contributes to the improved quality of telephone interviewing. Due to technical reasons, this could not be ascertained for the present data. However, even if this is so, it does not change the basic conclusions, but, possibly, provides a partial explanation.

To summarize, its limitations notwithstanding, the misclassification model indicates systematic significant differences between responses obtained for the same units by home visits and by telephone interviews. A comparison for simple response variances has led O'Muircheartaigh (1986) to conclude that telephone interviewing is more reliable than personal interviewing. Although the differences are not substantial, our results also clearly indicate that inconsistency in response – as measured by the rate of misclassification for qualitative variables – is higher for responses obtained by home visits than for those obtained by telephone interviews and that biases in estimating a distribution can be reduced by the use of telephone surveying.

5. References

Bailar, B.A. (1975): The Effects of Rotation Group Bias on Estimates from Panel Surveys. Journal of the American Statistical Association, 70, pp. 23–30.

Central Bureau of Statistics (1985): Housing Conditions and Possession of Household Equipment. 1983 Census of Population and Housing, Publication No. 9, Jerusalem.

Central Bureau of Statistics (1987): Labour Force Surveys 1985. Special Series No. 801, Jerusalem.

Christoffersen, M. (1984): The Quality of Data Collected at Telephone Interviews. Statistisk tidskrift, 22, (1), pp. 27–35.

Collins, M. (1983): Telephone Interviewing in Consumer Surveys. Market Research Society Newsletter, No. 11.

Colombotos, J. (1969): Personal Versus Telephone Interviews; Effect on Responses. Public Health Reports, 84:9, pp. 773–782.

Cowan, C.D. (1985): Misclassification of Categorical Data. American Statistical Association, Proceedings of the Social Statistics Section, pp. 491–496.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977): Maximum Likelihood from Incomplete Data Via the EM Algorithm. Journal of the Royal Statistical Society, Series B, 39, pp. 1–38.

Groves, R.M. and Kahn, R.L. (1979): Surveys by Telephone: A National Comparison with Personal Interviews. Academic Press, New York.

Hochberg, Y. (1977): On the Use of Double Sampling Schemes in Analyzing Categorical Data with Misclassification Errors. Journal of the American Statistical Association, 72, pp. 914–921.

Hochstim, J.R. (1967): A Critical Comparison of Three Strategies of Collecting Data From Households. Journal of the American Statistical Association, 62, pp. 976–989.

Kantorowitz, M. and Nathan, G. (1987): The Estimation of Response Error Micro-effects From Repeated Surveys for Invariant Characteristics. Proceedings of the Third Annual Research Conference, Bureau of the Census, Washington, D.C., pp. 359–390.

Klecka, W.R. and Tuchfarber, A.J. (1978): Random Digit Dialing: A Comparison to Personal Surveys. Public Opinion Quarterly, 42, pp. 105–114.

Locander, W.B., Sudman, S., and Bradburn, N.M. (1974): An Investigation of Interviewing Method, Threat and Response Distortion. American Statistical Association, Proceedings of the Social Statistics Section, pp. 21–27.

Nathan, G. (1987): The Use of Linear Models and Misclassification Models to Assess Response Error Effects of Mode of Data Collection. Bulletin of the International Statistical Institute, 52, 13.2.

Nicholls, W.L. and Groves, R.M. (1985): The Status of Computer Assisted Telephone Interviewing, 1985. Bulletin of the International Statistical Institute, 51, 18.1.

O'Muircheartaigh, C. (1986): Correlates of Reinterview Response Inconsistency in the Current Population Survey (CPS). Proceedings of the Second Annual Research

Conference, Bureau of the Census, Washington D.C., pp. 208–234.

Press, S.J. (1968): Estimating from Misclassified Data. Journal of the American Statistical Association, 63, pp. 123–133.

Rogers, T.F. (1976): Interviews by Telephone and in Person: Quality of Responses and Field Performance. Public Opinion Quarterly, 40, pp. 51–65.

Schuman, H. and Presser, S. (1981): Questions and Answers in Attitude Surveys. Academic Press, New York.

Sykes, W. and Hoinville, G. (1985): Telephone Interviewing on a Survey of Social Attitudes: A Comparison with Face to Face Procedures. SPCR Survey Research Centre, London.

Waksberg, J. (1978): Sampling Methods for Random Digit Dialing. Journal of the American Statistical Association, 73, pp. 40–46.

Wiseman, F. (1972): Methodological Bias in Public Opinion Surveys. Public Opinion Quarterly, 36, pp. 105–108.

Received December 1987
Revised December 1988