THE PHILIPPINE STATISTICIAN
Volume XVII, Nos. 3/4, 1968

SOME NON-SAMPLING ERRORS IN SAMPLE SURVEYS[1]

by

C. P. PAREL[2]

1. Introduction:

The main interest in sample surveys is the characterization of the populations from which these samples are taken. These characterizations are usually expressed as estimates of the population characteristics. But these estimates are not without error as they are made from data collected from selected sample from the population.

In any sample survey, the accuracy of the collected data is affected by two kinds of errors: *sampling errors*, which arise due to the non-coverage of the entire population under investigation, and the *non-sampling errors*, which are present whether or not a complete coverage of the population is made. As the term "non-sampling" connotes, these errors are due to factors other than generalizing from a sample to the population. There are several sources of non-sampling errors. Some of them are: the interviewer due to inappropriate methods of interviewing, or inaccurate recording of responses, the inaccurate responses of the respondent, non-response, incorrect processing of basic data and others. In this study, "non-response" is where an individual is included in the sample but could not be interviewed after a number of "call-backs".

Much have been written on the measurement and control of sampling errors, but very little has been done on the

[1] The study was financed jointly by the U.P. Statistical Center and the Social Science Research Council.
[2] Director, The Statistical Center, U.P.

measurement and control of non-sampling errors. Various ingenuities have been shown in devising techniques for the assessment and control of non-sampling errors. This is a field on which broad generalizations are difficult to attain. However, with the growing experience with sample surveys, information should gradually accumulate about the nature and magnitude of non-sampling errors particularly that of response in different types of surveys.

In view of the increasing popularity in the Philippines of sample surveys as a tool in research and the lack of studies on the assessment and control of non-sampling errors, it was decided to design a study of the errors arising from the interviewers, non-response, and the respondent themselves.

The present study is merely a first attempt to study errors in survey results arising from data collection. It is intended to be a "pilot study" to get an idea of how future studies on non-sampling errors may be designed and to gain insight into the problems that studies of this nature may encounter.

2. Objectives:

Using household surveys the study primarily attempted:

i) To determine the characteristics of non-respondent households;

ii) To ascertain the response rate of households with varying characteristics to certain types of questions;

iii) To discover methods of maximizing possible response rates in order to minimize non-response bias;

iv) To estimate the magnitude of response errors by types of questions;

v) To ascertain differential response rates for urban and rural respondents; and

vi) To ascertain response differences due to sex and length of training of the interviewers.

3. **Design of the Study:**

The major sources of variation that were studied in the present investigation are: the sex of the interviewers, the length of training of the interviewers and the respondent households.

The population of households was classified according to size of households, educational attainment of head of household and occupation of the head of household. There were 4 x 2 x 3 or 24 strata of households formed according to the following categories:

i) Occupation of head of household (O):

O_1 — students only, jobless, housewives;
O_2 — daily wage earners, on commission basis, small shop owners;
O_3 — regular employees pensioners;
O_4 — executives, landowners, realtors.

ii) Size of household (H):

H_1 — 4 and below;
H_2 — 5 and above.

iii) Educational attainment of head of household (E):

E_1 — below high school
E_2 — high school
E_3 — college and above.

Two areas in Greater Manila (Paco and Kamuning) and one barrio in Bulacan (Bambang) were chosen for the study. Paco was chosen in the city because of its proximity to the Statistical Center, while Kamuning was chosen because it was believed that the people living in the area are typically of the middle income class. Bambang, a barrio of Bulacan, was selected as the rural area for study not only for convenience (as it is near Manila), but also because the majority of the people there thrive on rice farming and fishing.

For both Paco and Kamuning, two independent samples of 8 households were randomly chosen (without replacement) from each of the 24 groups of households. If the group contained less than 8 households, all were taken. In Paco, there were 188 sample households for each of the two phases; there were 182 sample households in Kamuning. One sample was used in the first interview while the second was used in the second interview which was conducted about a month after the first. Since all the households selected for the first sample were replaced before the choice of the second sample, some households were common to both samples.

The chosen households in each group were randomly assigned to 4 interviewers for both Paco and Kamuning: two males and two females, all of whom have had four years of college. Two of the interviewers (one male and one female) were trained for two days, while the other two (one male and one female) were trained for four days.

Phase I—Interview	Phase II—Interview
Male - long training (ML)	Female - long training (FL)
Female - short training (FS)	Male - short training (MS)
Female - long training (FL)	Male - long training (ML)
Male - short training (MS)	Female - short training (FS)

The population of 917 households in Bambang, Bulacan, were classified into 4 x 2 or 8 groups according to the educational level of the head of household (primary, intermediate, high school, and college level) and household size (4 and below, and 5 and above). Two independent samples, each of size 194 households, were chosen from the population for the two interviews by randomly drawing without replacement about 25 households from each of the 8 strata. The 194 households were randomly distributed among 7 interviewers who are residents of the place (4 females and 3 males). Two females were given training for only two days; the other two females, 4 days. Two male interviewers were given training for 4 days; one male interviewer, only two days.

The schedule used in the rural area was slightly revised to suit the level and culture of the people in the area. The questions on current issues were modified and some of the questions under household expenditures were omitted. Questions eliciting farm information were included.

3. Collection of Data

The formal start of the interview was in the afternoon of February 18, 1967. Priority respondent in the sample households was the head of the household. If the head was not available, acceptable substitute was the wife of the head, if married. Where the household head or the wife were unavailable for interview, voluntary responses of another household member, a friend or neighbor were not accepted no matter how close this volunteer may be to the sampled household in question. The interviewer was asked to follow up said household three times, and if still unavailable after the third time, this particular household was considered a non-response. However, for purposes of the study, such households were replaced by other households chosen at random, and having the same economic characteristics, educational attainment and household size. Where there is no substitute available in the population of that particular cell or block of households, this household was just recorded either as a case of "refusal", "not available", or "transferred", as the case may be.

4. Definition of Terms:

In this study the following definitions were adopted.

Household — composed of persons who may or may not be related to each other but who sleep in the same dwelling unit and usually have a common arrangement for the preparation and consumption of food. (Bureau of Census and Statistics' definition.)

Household head — the bread winner (main support) of the family.

Occupational group — a group of occupations as classified by the Statistical Center for the purpose of the Survey. (Section 2).

Mother tongue — the dialect spoken in a person's home in his earliest childhood.

Non-coverage ratio — the ratio of the sample respondent households, who were not contacted, to the total number in the sample.

Non-response — the term used when information is not gathered from a sample respondent after a number of call backs.

Response error — a response different from that given during the listing.

Refusal — a respondent that was not interviewed after 3 callbacks.

Pair of schedules — schedules used for the same sample household in two interviews.

5. Problems Encountered in the Study:

In view of the varied problems encountered in the field, the original design was not strictly followed. For example, originally, the design included the "time of interview" as one of the variables under study, but this variable had to be ignored when the interviewers encountered all sort of problems in the field. They could not only locate their respondents but it was difficult for the interviewers to "catch" them even on Saturdays and Sundays. For this reason, the interviewers were instructed to make the interview whenever they could.

It was decided to wait for about a month after the listing of households in Paco and Kamuning before the interviewers went back to the field to interview the sample respondents. It was believed that the sample households might not cooperate with the survey if they were to be approached again soon after the listing. But when the interviewers went back to the field, a number of sample households could no longer be located for most of them had moved to other places. There was also that big conflagration in Paco a week after the interviewers

were "fielded", so that sample households included in the fire area had to be replaced.

One of the male interviewers had to be replaced before the field work was completed because he could not keep up with his assigned schedule. This sort of disrupted the original design.

The major difficulty in the study, though, is the lack of information against which the reported responses could be checked. There were only three items which were collected during the listing which could be used as checks to the information gathered during the two interviews. The most that the study could do, therefore, was to compare and determine the extent of agreement between the responses of the same households that happened to be included in the two independent samples used for the two interviews. In this comparison, it was difficult to ascertain, however, who between the two, the respondent or the interviewer, is the source of the error.

Information about the household head, or the household was intended to be collected. It appeared, however, in some interviews that the respondent was not only the head of the household, but any adult member of the household that happened to be available. This accounted for some of the discrepancies found in the reported responses.

Due to the above difficulties and problems met in the study, the results of this survey are by no means conclusive. Further research is planned in this field, whereby, it is hoped, some of the difficulties could be avoided.

6. Summary of Findings and Recommendations:

1. In Paco about 59% of the submitted schedules in the two interviews did not agree with the information elicited during the listing (viz. head of household, size of household, and occupation of head of household). The corresponding pro-

portion in Kamuning is about 58%. In Bambang, the rural area, the responses given on those three items during the listing and actual interview tallied completely. Many of the discrepancies were due to the fact that different respondents gave the information during the listing and actual interviews.

TABLE 1. DISCREPANCIES OF RESPONSES DURING LISTING AND INTERVIEW BY TYPE AND AREA

Type	Phase I				Phase II			
	Paco		Kamuning		Paco		Kamuning	
	No.	%	No.	%	No.	%	No.	%
Total	183[1]	100.0	170[1]	100.0	181[1]	100.0	172[1]	100.0
No change	59	32.24	77	45.30	84	46.41	68	39.53
Change in occupation only	58	31.69	30	17.65	42	23.21	44	25.58
Change in size of household only	30	16.38	30	17.65	22	12.15	27	15.70
Change in head of household only	2	1.10	2	1.18	1	0.55	2	1.16
Change in occupation & size of household	18	9.84	11	6.47	14	7.74	19	11.05
Change in occupation & head of household	5	2.73	9	5.29	10	5.52	5	2.91
Change in size and head of household	4	2.20	2	1.18	—	—	—	—
Change in occupation, size and head of household	7	3.82	9	5.28	8	4.42	7	4.07

[1] These include the respondents that were used to replace those that could not be covered after 3 call-backs.

2. The study seems to indicate that the number of errors is independent of the sex of the interviewers. The length of training does not seem to matter very much when quantitative information is to be collected.

3. The "non-coverage" ratios in Paco and Kamuning are about 25% and 15% respectively. About 14% were "non-responses" in Paco; about 10% in Kamuning. In Bambang, the "non-coverage" ratio is only about 6%, all of which were "non-

responses". It seems, therefore, that the cooperation of households in rural areas is easier to enlist in field surveys than those in the urban areas.

TABLE 2. NUMBER OF COMPLETE NON-RESPONSES CLASSIFIED BY REASONS FOR NON-RESPONSE AND BY TYPE OF INTERVIEWER

A. PACO

Type of Interviewer	PHASE I					PHASE II				
	Reasons			Total		Reasons			Total	
	R	NA	T	No.	%	R	NA	T	No.	%
Total	13	8	11	32	100.00	6	25	32	63	100.00
MS	8	2	4	14	43.75	2	7	11	20	31.75
ML	1	—	4	5	15.63	3	3	11	17	26.98
FS	4	5	1	10	31.25	1	7	4	12	19.05
FL	—	1	2	3	9.37	—	8	6	14	22.22

B. KAMUNING

Total	7	14	8	29	100.00	5	13	9	27	100.00
MS	—	4	2	6	20.69	—	2	—	2	7.41
ML	5	2	1	8	27.59	4	3	1	8	29.63
FS	1	4	3	8	27.59	—	2	2	4	14.81
FL	1	4	2	7	24.13	1	6	6	13	48.15

C. BAMBANG, BULACAN

Total	4	1	—	5	100.00	9	7	—	16	100.00
MS (1 int.)	1	—	—	1	20.00	2	—	—	2	12.50
ML (2 ints.)	2	—	—	2	40.00	4	1	—	5	31.25
FS (2 ints.)	1	1	—	2	40.00	1	—	—	1	6.25
FL (2 ints.)	—	—	—	—	——	2	6	—	8	50.00

Note: R—Refused; NA—Not Available; T—Transferred

4. No general conclusions can be made about the characteristics of households, the cooperation of which is difficult to enlist in household surveys. In Paco, the households that belong to the highest socio-economic group appear to be the most non-cooperative judging from the proportion of non-responses; but in Kamuning, this does not seem to be the case.

Also, in Paco, the refusals seem to decrease with higher educational attainment; in Kamuning, there were more refusals among the more highly educated group.

For households which belong to those above groups, (those in high socio-economic groups), the survey leader or the most senior interviewer should be assigned to conduct the interview. The supervisor was able to interview those households in Paco that refused to cooperate previously.

5. The female interviewers seem to be better in getting the cooperation of the respondents than the male interviewers. However, the male interviewers seem to be better in "tracking down" the respondents than the female interviewers. A number of these households reported by the female interviewers to be not "available" for interview live in somewhat isolated places.

Hence, in "difficult places" — places that are some distance from bus-lines or public conveyances — the male interviewers should be assigned.

TABLE 3. NUMBER OF UNANSWERED ITEMS IN INCOMPLETE SCHEDULES CLASSIFIED BY SEX AND LENGTH OF TRAINING OF INTERVIEWERS (PHASES I & II)

Sex of Interviewer	PACO		KAMUNING		BAMBANG, BULACAN	
	No	%	No.	%	No.	%
Total	524	100.00	145	100.00	577	100.00
Female	319	60.88	50	34.48	257	44.54
Male	205	39.12	95	65.52	320	55.46
Length of Training	**No.**	**%**	**No.**	**%**	**No.**	**%**
Total	524	100.00	145	100.00	577	100.00
Short	247	47.13	72	49.66	381	66.03
Long	277	52.87	73	50.34	196	33.97

6. In all the places covered in the study, the questions on "attitudes to current issues" and "exposure to mass media and community participation" were the most frequently left unanswered. These items consisted of some open-ended ques-

tions. In the rural area covered, quite a number did not answer those whose reasons for the given responses were asked. About one out of two respondents in the rural area did not respond completely to the open-ended questions; it was one out of three in Paco and one out of ten in Kamuning. There seems to be evidence for urban people to be more "talkative" than the rural people.

It seems necessary, therefore, that when there are open-ended questions in the schedule, or when the respondents are supposed to elaborate on their answers, interviewers who are naturally good "talkers" and can get the respondents to talk, should be assigned.

7. In repeated interviews, almost all of the matched or paired-schedules in the two urban areas did not completely agree. In the rural area, about 11% of the matched schedules did not agree completely.

This finding is a very disturbing one especially for researchers who depend much on data gathered in the field. It seems necessary, therefore, in order to ascertain how much confidence to be placed in survey results, to "build-in" the design a sort of a check of the responses after the actual interview. This can be done by taking a small sub-sample of the original sample for re-interview to ascertain the reliability of the responses given.

8. The low correlation coefficients between the reported average total monthly incomes in the interviews in Paco ($r = 0.4219$) and Kamuning ($r = 0.1989$) point to the difficulty of getting correct information on household income in household surveys. A more "correct" information on income seems to be possible from the rural areas ($r = 0.9670$). The same observation applies to the collection of information on average monthly rice expenditure and on age.

9. There is no "best" time to interview household respondents. What is "best" time for one respondent may not be for another. The "best" time to interview sample respondents is that time when the interviewer can "catch" them and these respondents are ready to cooperate.

TABLE 4A. NUMBER OF INCOMPLETE SCHEDULES BY ITEM OF INFORMATION NOT ANSWERED AND BY TYPE OF HOUSEHOLDS

A. PACO

Item of Information	Occupational Groups	PHASE I						Total
		Size of Household						
		4 and Below			5 and Above			
		Educational Level of Head			Educational Level of Head			
		Below H. S.	H.S.	College and Above	Below H. S.	H.S.	College and Above	
	Total	20	12	11	23	24	24	114
Demographic characteristics	O_1	2	—	—	—	—	—	2
	O_2-O_4	—	—	—	—	—	—	—
Income	O_1	2	—	—	—	—	—	2
	O_2	—	—	—	—	—	—	—
	O_3	—	—	—	—	1	2	3
	O_4	—	—	—	—	—	—	—
Household expenditures	O_1	1	1	1	1	—	—	4
	O_2	—	1	1	—	—	4	6
	O_3	1	4	—	—	4	1	10
	O_4	2	—	—	1	1	1	5
Exposure to mass media and community participation	O_1	1	—	—	—	—	—	1
	O_2	—	—	1	2	2	—	5
	O_3	1	—	2	1	1	1	6
	O_4	—	—	—	—	—	—	—
Attitude to current issues	O_1	2	—	2	1	2	5	12
	O_2	2	3	3	4	3	6	21
	O_3	5	3	1	9	10	3	31
	O_4	1	—	—	4	—	1	6

TABLE 4A. NUMBER OF INCOMPLETE SCHEDULES BY ITEM OF INFORMATION NOT ANSWERED AND BY TYPE OF HOUSEHOLD — (CONTINUED)

A. PACO

Item of Information	Occupational Groups	PHASE II						Total
		Size of Household						
		4 and Below			5 and Above			
		Educational Level of Head			Educational Level of Head			
		Below H. S.	H.S.	College and Above	Below H. S.	H.S.	College and Above	
Demographic characteristics	Total	15	23	12	10	16	9	85
	0_1-0_4	—	—	—	—	—	—	—
Income	0_1	—	—	—	1	2	—	3
	0_2	1	—	1	1	—	1	4
	0_3	2	2	1	—	—	2	7
	0_4	—	4	2	—	1	—	7
Household expenditures	0_1-0_4	—	—	—	—	—	—	—
Exposure to mass media and community participation	0_1	—	—	—	—	1	—	1
	0_2	2	1	—	—	1	1	5
	0_3	—	1	—	1	1	—	3
	0_4	—	2	1	—	1	—	4
Attitude to current issues	0_1	—	2	—	1	2	1	6
	0_2	4	1	2	3	1	3	14
	0_3	4	2	2	2	3	1	14
	0_4	2	8	3	1	3	—	17

TABLE 4A. NUMBER OF INCOMPLETE SCHEDULES BY ITEM OF INFORMATION NOT ANSWERED AND BY TYPE OF HOUSEHOLDS — (CONTINUED)

B. KAMUNING

Item of Information	Occupational Groups	PHASE I						Total
		Size of Household						
		4 and Below			5 and Above			
		Educational Level of Head			Educational Level of Head			
		Below H. S.	H.S.	College and Above	Below H. S.	H.S.	College and Above	
	Total	4	6	9	4	4	3	30
Demographic characteristics	O_1-O_4	—	—	—	—	—	—	—
Income	O_1-O_3	—	—	—	—	—	—	—
	O_4	—	—	—	—	—	1	1
Household expenditures	O_1	—	—	1	—	—	—	1
	O_2-O_3	—	—	—	—	—	—	—
	O_4	—	—	1	—	—	—	1
Exposure to mass media and community participation	O_1	—	—	1	—	—	1	2
	O_2	2	1	—	1	1	—	5
	O_3	—	2	—	—	—	—	2
	O_4	—	—	1	2	1	—	4
Attitude to current issues	O_1	—	—	1	—	—	1	2
	O_2	1	2	—	—	2	—	5
	O_3	1	1	2	—	—	—	4
	O_4	—	—	2	1	—	—	3

TABLE 4A. NUMBER OF INCOMPLETE SCHEDULES BY ITEM OF INFORMATION NOT ANSWERED AND BY TYPE OF HOUSEHOLD — (CONTINUED)

B. KAMUNING

Item of Information	Groups Occupational	PHASE II						Total
		Size of Household						
		4 and Below			5 and Above			
		Educational Level of Head			Educational Level of Head			
		Below H. S.	H.S.	College and Above	Below H. S.	H.S.	College and Above	
	Total	3	12	1	6	4	10	34
Demographic characteristics	0_1-0_4	—	—	—	—	—	—	—
Income	0_1-0_2	—	—	—	—	—	—	—
	0_3	—	—	1	—	1	—	2
	0_4	—	2	—	1	—	—	3
Household expenditures	0_1-0_4	—	—	—	—	—	—	—
Exposure to mass media and community participation	0_1	—	1	—	—	—	—	1
	0_2	—	—	—	1	—	2	3
	0_3	—	—	—	—	—	2	2
	0_4	—	2	—	—	—	—	2
Attitude to current issues	0_1	1	1	—	—	—	—	2
	0_2	1	1	—	2	—	2	6
	0_3	1	1	—	1	2	2	7
	0_4	—	4	—	1	1	—	6

TABLE 4B. SUMMARY OF INCOMPLETE SCHEDULES CLASSIFIED BY ITEMS OF INFORMATION AND BY EDUCATIONAL LEVEL AND SIZE OF HOUSEHOLD (BAMBANG, BULACAN)

Items of Information	PHASE I									PHASE II								
	Primary		Inter-mediate		High School		College		Total	Primary		Inter-mediate		High School		College		Total
	4 and Less	5 and Over	4 and Less	5 and Over	4 and Less	5 and Over	4 and Less	5 and Over		4 Less and	5 Over and	4 Less and	5 Over and	4 Less and	5 Over and	4 Less and	5 Over and	
To t a l	23	25	25	21	14	21	14	17	160	16	33	23	21	19	26	16	17	171
Demographic characteristics	5	4	3	1	2	2	3	5	25	—	—	1	1	2	3	1	3	11
I n c o m e	—	—	—	—	—	—	—	—	—	—	1	1	—	1	1	2	—	6
Household expenditures	—	1	1	—	—	1	—	1	4	—	—	—	1	1	—	1	—	3
Exposure to mass media and community participation	7	5	5	9	3	5	2	2	38	6	12	7	7	7	11	5	4	59
Attitudes on current events	9	13	15	10	9	12	7	8	83	9	20	14	12	8	11	7	10	91
Farm information	2	2	1	1	—	1	2	1	10	1	—	—	—	—	—	—	—	1