INTERVIEWS HARDEST-TO-OBTAIN IN AN URBAN HEALTH SURVEY

Regina Loewenstein, John Colombotos, Jack Elinson Columbia University School of Public Health and Administrative Medicine

Introduction

Organizations doing interview surveys are always concerned about the possible bias in their results due to the persons in their samples who were not interviewed. For example, the proportion of persons with a particular socio-economic characteristic may be over or under represented among those interviewed. Statisticians are asked to measure the bias due to non-response. Supervisors of field operations would like to know the types of persons who are hardest-to-interview in order to develop techniques to interview them.

These questions were the concern of the staff of a collaborative research project called the Master Sample Survey being carried out by the Columbia University School of Public Health and Administrative Medicine in the Washington Heights Health District in New York City<u>1</u>/•

The Master Sample was designed in 1959 by the Columbia-Washington Heights Community Mental Health Project so that random samples or groups with specified demographic characteristics could be selected for small studies. By early 1960, the plan was to use this sample for a collaborative research project on a variety of medical and health problems. This work was supported by the Health Research Council of the City of New York, under Contract U-1053.

The five research groups participating in the 1960-1961 Master Sample Survey were: Patterns of Medical Care Study of New York City Department of Health, Columbia-Washington Heights Mental Health Project, National Council on Alcoholism, and Cardiovascular and Neurological study groups of the Columbia University School of Public Health and Administrative Medicine.

During 1960, interview schedules were designed by John Colombotos in collaboration with representatives of each of these research groups to answer questions of interest to one or more of them.

From November 1960 through April 1961, interviews were sought from a two-stage stratified cluster sample with 4,500 dwelling units2/. By the first cut-off date, interviews had been obtained from 3,300 dwelling units or 77 percent of eligible dwelling units, allowing for sampling losses.

The staff wanted to study the possible bias due to the absence of interviews with about onefourth or 1,000 of the dwelling units in the sample on the first cut-off date. Persons in 550 of these dwelling units had been contacted and refused to be interviewed. The remaining 450 dwelling units had not been interviewed for reasons other than refusal; namely, no contact had been made or some contact had been made but with no direct refusal.

To study the characteristics of this "hard-to-obtain" group, a sample of one-third or 340 of the dwelling units with no interviews were selected for intensive investigation. Between May 1st and July 15th, members of the supervisory field staff, two medical students2/, and several of the best interviewers exerted intensive effort to obtain interviews in these 340 dwelling units.

By the final cut-off date⁴/, interviews had been obtained in 210, or about 60 percent of the dwelling units. About one-half of the previous refusals and three-fourths of the previous nonrefusals had been interviewed. Of the remaining 130 dwelling units with no interviews, almost 80 percent had refused at least once, and only 20 percent had never been contacted. Thus, persons never interviewed were primarily people who had refused to be interviewed at least once.

The 3,300 early interviews gave information about 3,700 families 2/and 9,200 persons. Adding the interviews obtained in the second round, weighted by three, the tabulations of this survey are based on 4,000 interviews with 4,400 families and 10,750 persons. Data about the 4,400 families will be referred to as marginal distributions.

In Table 1 are the distributions of size of family and age of persons of the early and late cases, with the late cases classified as previous refusals and previous non-refusals according to their status as of the first cut-off date. The distributions of these two characteristics are almost identical for the Washington Heights Health District, the estimated sample sought, the early cases, and the marginal distributions.

There are differences in size of family and age of persons between the early and late cases. In spite of this, the marginal distributions and the distributions of the early cases are almost identical because, even after weighting, the 229 families in late cases are only 16 percent of the 4,400 families represented in the marginal distributions.

NOT INTERVIEWED GROUP

Since the survey had been done within one year of the 1960 United States Census of Population, tabulations were purchased from the Bureau of the Census on a few characteristics about persons in dwelling units with no interviews for whom names were known provided that the names and addresses were exactly the same at the time of the Census. Tables about 94 matched dwelling units gave data on size of family, sex, race and age, of about 200 persons living in these dwelling units.

It was then possible to estimate the distributions of these 4 characteristics for 98 percent of the dwelling units eligible for interview from the sum of (a) the early cases; (b) the late cases, multiplied by 3; and (c) the non-interviews in the one-third sample, multiplied by 3. The characteristics of the remaining 2 percent, that is, dwelling units for which no names were known plus those whose names were not found at the specified address in the Gensus records, were assumed to have the same characteristics as non-interviews that had been found.

Table 2 summarizes the results of this procedure. The sample sought was estimated to be 4,800 families with 11,500 persons. At the first cut-off date, 77 percent of the families and 80 percent of the persons were represented in the interviews. After weighting the late interviews by 3; 91 percent of the families and 93 percent of persons in the sample were represented.

By the first cut-off date, interviews had been obtained from 73 percent of families of one and two persons, and 83 percent of families of three or more persons. During the second round, the interviewing rate among larger families remained higher, and interviews obtained by the second cut-off date represented 89 percent of families of one or two persons and 96 percent of families of three or more persons.

There were no differences in the proportions of men and women or of white and nonwhite persons who were interviewed early or late, or who did or did not refuse to be interviewed in the first round.

Table 2 also shows differences in ease of interviewing persons in various age groups. By the first cut-off date, 85 percent of persons under 45 years, and 75 percent of persons 45 and older had been interviewed. During the first round, the refusal rate had been highest among persons 65 years and older. In the second round the refusal rate was highest among persons 45 to 64. After the second cut-off date, 97 percent of persons under 45, and 90 percent of persons 45 and older were represented.

The relationship of age, size of family, and field status have not yet been analyzed.

COMPARISONS OF GROUPS BY FIELD STATUS

Information about non-response bias can be obtained by comparing early and late interviews, and also by comparing previous refusals and previous non-refusals, with regard to various characteristics. Such comparisons have been done for about 200 variables, and the results of these analyses will be summarized. All differences mentioned are significant on the five percent level. Significantly different percentages usually differ absolutely by 5 to 15 percent.

1. <u>Demographic and Socio-Economic Char-</u> acteristics.

The distributions of the early and late cases and of previous refusals and non-refusals with regard to size of family and age of persons are shown in Table 1.

Families who had refused to be interviewed in the first round included more two-person families than did the early cases and the previous non-refusals. Among the previous non-refusals were more one-person families, reflecting the difficulties interviewers had to find them at home.

Compared with the early cases, the previous non-refusals had more persons 45 to 64 years of age, and the previous refusals had more persons 65 and older.

There were no differences between early and late cases, or between previous refusals and nonrefusals with regard to race, sex, religion, rent, family income, public assistance, occupations of employed persons, countries of origin of heads of families, plans to move, reporting no close relatives, or having a family doctor.

Many of the characteristics that had different distributions among early and late cases may be related to the larger proportions of small families or older persons in the late cases. The late cases had more families with no children under 16, with no persons under 35, and at least one person 65 or older; and more widowed persons. Fewer of the adults in the late cases were high school graduates. Re occupation and income the employed persons in the late cases were more frequently self-employed, earning at least \$6,000 a year, and on the same job for at least 10 years. Fewer adults in the late cases worked full-time. Re previous residences - the late cases had more persons born in the United States outside of New York City, and more families who had been at the same address for at least 15 years.

Some of the characteristics which were different in the previous refusals and previous nonrefusals may be related to the differences in family size and age. The previous refusals had more families with all persons over 65 years, more adults who kept house, fewer adults who worked full time, and more adults with no high school education. The health insurance coverage was less comprehensive for the persons in the previous refusals.

Other characteristics which differed among previous refusals and non-refusals may be related to the larger proportion of persons born in Western Europe among previous refusals. While the field work was in progress, the interviewers reported difficulties in obtaining interviews with some refugees from Nazi governments because of their fear to let strangers into their homes. Among persons not born in the United States, the percentage who became citizens was higher among the previous refusals than the previous nonrefusals. Also, the previous refusals had more heads of families who had lived in New York City for at least 15 years. More previous refusals reported that their closest relative was other than a parent, child, or sibling.

2. <u>Reported chronic conditions, etc.</u>

Early and late cases, and previous refusals and non-refusals did not differ with regard to the reported presence during the past year of 29 chronic conditions, 12 out of 13 impairments, 4 symptoms of cardiovascular conditions, drinking problems, institutional care, or limitations in ordinary activities due to illness.

These findings seemed surprising because there were more older persons in the late than early cases, and in the previous refusals than the previous non-refusals. Therefore, the proportions of persons in 12 age-sex groups reporting at least one of the 29 chronic conditions were analyzed. In 10 of the 12 groups, the percents of persons in late cases reporting at least one condition were lower than among the early cases, and in 9 of the 12 groups the percents were lower among previous refusals than previous non-refusals.

Were persons in the late cases, especially previous refusals, healthier or more laconic in the interview?

3. Reported medical care during past year.

There were no significant differences between early and late cases in the proportions of persons reporting attended conditions, the mean number of reported visits per person interviewed or per person reporting at least one attended condition.

But a significantly smaller proportion of the previous refusals reported one or more attended conditions. The mean numbers of reported visits to each type of facility were smaller for the previous refusals than the nonrefusals, but the differences were not significant.

The late cases and the previous refusals had fewer persons with at least one hospital stay, and more of the stays reported by the late cases were 10 days or more. The reported numbers of days in bed for reported attended conditions were the same for persons in the three groups of cases, but the previous non-refusals reported more persons than did the previous refusals who missed daily activities 20 or more days because of reported attended conditions. There were fewer persons among the previous refusals reporting 10 or more out-of-hospital visits in one year than among the early cases or among the previous non-refusals.

The proportions of persons reporting that all out-of-hospital care was from doctors in private practice were the same among the early cases, previous refusals, and previous nonrefusals. But previous refusals went more often than previous non-refusals to doctors whom they called "family doctors." The previous non-refusals had more visits to out-patient departments and more frequently used three or more different facilities for out-of-hospital care in one year.

Among the previous refusals, in comparison with previous non-refusals, there were more reported attended conditions with the first and only care from a doctor known before, and fewer with any care from a facility referred by a medical person.

Further analyses may help in the interpretation of the differences in reported chronic and attended conditions for persons in the previous refusal and non-refusal cases. How are these differences related to perceived need for medical care, the amount and type of medical care sought and received, and the cooperation of interviewees in the two groups of cases?

4. Questions asked of Sample of Adults.

Questions suggested by the Patterns of Medical Care Study and the Columbia-Washington Heights Community Mental Health Project were asked of a sample of persons 21 years and older in the Patterns Cases.

Marvin Belkin of the staff of the Patterns of Medical Care Study developed 11 indices based on their responses to 37 multiple choice questions. There were no significant differences between early and late cases or between previous refusals and previous non-refusals on seven indices; namely, Awareness of Physician's Social Status, Physician's Interest in Patient's Welfare, Acceptance of Sick Role, Preventive Medical Behavior, Illness-Dependence Orientation, Work Indispensability, and Ethnic Exclusivity.

The late cases had more persons with low indices re Reliance on Friends and also Friendship Solidarity, and previous non-refusals had more with low indices on Friendship Solidarity. On the Index for Family Orientation to Tradition and Authority, the late cases and the previous refusals both had more with high scores.

An index on Attitude toward Medical Care was based on responses to three questions about trying different doctors, doubting doctors can help, and wanting details when ill. These questions were:

were: "I believe in trying different doctors to find which one I think will give me the best care."

"When I am ill, I demand to know all the

details of what is being done to me."

"I have my doubts about some things doctors say they can do for me."

Persons were asked whether they agreed or disagreed with these statements, and could also give qualified responses.

There were more persons among the previous non-refusals than previous refusals who agreed with all three of these questions, which was classified as lowest on this index. Further analyses may show if a low index on Attitude toward Medical Care is related to more reported medical care.

There were no differences by previous field status as to the reporting of dental care, eye examinations, periodic checkups, polio shots, balanced diet, or expressed need for additional health services. But, the early cases had a higher score on the Knowledge of Illness test than the late cases, possibly related to the differences in education.

Bruce Dohrenwend of the Columbia Washington Heights Mental Health Project suggested the inclusion of 22 questions from the Midtown Manhattan Study^O. These items were found to discriminate between psychiatric patients and well persons^I. The items used were originally drawn from the neuropsychiatric screening inventory used during World War II and the Minnesota Multiphasic Personality Inventory. A score was developed from the responses to these 22 questions. There were no differences between persons in early and late cases, or in previous refusals and non-refusals as to the distributions of this score or the average score.

The four field status groups also did not differ in responses to five questions on anger, guilt, and fear.

5. Note re Comparisons.

The previous refusals and previous nonrefusals differed from each other on more of the 200 characteristics than did the early and the late cases, because there were many characteristics in which the previous non-refusals were more like early cases than like the previous refusals.

Almost all of the comparisons that have been discussed are first order-relationships, that is, the two pairs of field status groups , as a whole have been compared without controlling for other variables. Exceptions were the comparisons of previous refusals and previous nonrefusals as to reported chronic conditions by age. Other analyses of second and third order relationships might give further understanding about the types of persons who refused to be interviewed in this health survey, or who were hard to interview for other reasons.

EARLY CASES AND MARGINAL DISTRIBUTIONS

To measure the effect on the final results of the intensive effort of the second round, comparisons with regard to these same 200 variables were made of the distributions of early cases and the marginal distributions; that is, early cases plus three times the late cases.

The distributions of 135 variables differed absolutely by less than one percent in all categories, and 50 variables differed by one to two percent in one category. In the 15 variables with absolute differences of two to three percent, there were usually small bases for the percentages and/or large proportions of no answers to specific questions among the late cases.

That is, the distributions of very few characteristics would have been different if field work had stopped at the first cut-off date, and there had not been intensive effort to obtain interviews from a sample of the non-interviews at that time.

SUMMARY OF FINDINGS

1. By the first cut-off date, interviews had been obtained from 77 percent of 4,800 families; 12 percent had refused to be interviewed and 11 percent had not been interviewed for other reasons.

2. Intensive field effort on a one-third sample of cases not interviewed by the first cutoff date resulted in estimated response rates of 91 percent of families and 93 percent of persons eligible for interviews.

3. Interviews with families of only one and two persons were harder to obtain than interviews with larger families. Interviews with older persons, i.e., 45 years or older, were harder to obtain than interviews with younger persons. Further analyses are needed about the correlation of these findings.

4. Many of the socio-economic characteristics that differed among early and late interviewed cases appear to be related to age or family size. The early and late cases, however, reported the same amount of chronic and attended conditions.

5. One group of persons who were interviewed late had previously refused to be interviewed. This group of "previous refusals" differed from others who were interviewed late in a number of socio-economic characteristics. Some, but not all, of these differential characteristics were related to age or family size. Previous refusals reported fewer chronic and attended conditions than previous non-refusals of the same age.

6. Previous refusals and previous non-refusals did not differ in eight attitudinal areas. Previous refusals had lower indices on Friendship Solidarity, higher on Family Orientation toward Tradition, and a different Attitude about Medical Care.

7. There were no differences in the distributions of about 200 characteristics between early cases and final marginal distributions, i.e., early cases plus late cases weighted by three. For 135 of these variables the distributions never differed by more than one percent in any category, and there were never differences of more than three percent between the two distributions.

IMPLICATIONS

These findings, of course, would not necessarily be the same in other surveys with different interviewees, interviewers, and topics.

Frequent comparisons of the distributions of important characteristics while a survey is

- <u>1</u>/ According to the United States Censuses of Population and Housing, the Washington Heights Health District had 100,000 housing units and 270,000 persons in 1960.
- 2/ The sample was designed in 1958 by Marvin Glasser now at Harvard University.
- 3/ Charles Enzer and Donald Plevy, in cooperation with the 1961 Summer Training Program of the New York City Department of Health.
- 4/ Cut-off dates are in a sense arbitrary from a research point of view since they are influenced by "deadlines" for reports, time and personnel budgets, and, of course, available financial support.

in progress would inform the field staff what groups need more intensive effort. Decisions with regard to additional field work to be done would then depend on the nature and magnitude of the non-response bias, the available funds, and whether the major emphasis in the particular survey is to estimate the characteristics of a community or to study the inter-relationships of these characteristics.

In this survey, estimates of 200 social, health and medical care characteristics for the community as a whole from the first 77 percent of families in the sample were essentially identical with estimates based on interviews representing 91 percent of the families in the sample. Further analyses will show whether inter-relationships of these characteristics would have been effected if interviewing had stopped at the first cut-off date.

FOOTNOTES

- 5/ The number of families is estimated to be higher than the number of dwelling units because more than one family lived in some dwelling units.
- 6/ Leo Srole, Thomas S. Langner, Stanley T. Michael, Marvin K. Opler, Thomas A. C. Rennie, <u>Mental Health in the Metropolis</u>, New York: McGraw Hill, 1962.
- 2/ Thomas S. Langner, "Psychophysiological Symptoms and Women's Status in Two Mexican Communities," in Jane M. Murphy and Alexander H. Leighton (Editors), <u>Approaches to</u> <u>Cross Cultural Psychiatry</u>, in preparation.

TABLE 1

Size of Families and Age of Persons of the Washington Heights Health District, of the Estimated Sample, of Interviews received by Field Status, and Estimate of Non-interviews.

-				Interviews Received,				
• •					Late			[
	Washing- ton Heights Health Districtl	Esti- mated Sample	Margi- nal Distri- butions2/	Barly2/	Total Late	Pre- vious Refus- als	Pre- vious Non- Refus- als	Not Inter- viewed5/
Size of families Total number of families	100,987	4,804	4,392	3,705	229	104	125	
Percent distribution All families	100	100	100	100	100	100	100	100
1 person 2 persons 3 persons 4 or more persons	30 70	30 33 17 20	29 32 18 21	29 32 18 21	34 34 16 16	26 38 21 15	40 31 11 18	35 47 10 8
Age of persons Total number of persons	269,277	11,529	10,759	9,196	521	247	274	
Percent distribution All persons	100	100	100	100	100	100	100	100
1-15 years 15-44 years 45-64 years 65 yrs. and older	18 38 31 13	18 37 31 14	19 38 29 14	19 39 29 13	17 33 32 18	15 30 29 26	18 36 36 10	9 22 48 21

1/ Based on data by Census Tracts in United States Censuses of Population and Housing: 1960, Final Report PHC (1) - 104, Part 1.

2/ The results of the second phase of interviewing on a one-third sample of cases not completed in the first phase are weighted by 3. Marginal Distributions = Early Cases + 3 (Late Cases).

3/ Interviewed in November 1960 through April 1961.
4/ Interviewed in May through July 1961.
5/ Based on tabulations by the United States Bureau of the Census of a sample of not-interviewed cases matched with 1960 Census Records.

z.

			Interviewed Late			Not InterviewedZ/		
	Esti- mated Sample	Inter- viewed Early <u>1</u> /	Total Late	Pre- vious Refus- als	Pre- vious Non- Refus- als	Total Not Inter- viewed	Pre- vious Refus- als	Pre- vious Non- Refus- als
Dwelling units Estimated number	4,343	3,329	630	294	336	384	258	126
Percent distribution	100	77	14	6	8	9	6	3
<u>Size of families</u> Estimated number of families	4,803	3,705	687	312	375	411	273	138
Percent distributions All families	100	77	14	6	8	9	6	3
1 person 2 persons 3 persons 4 or more persons	100 100 100 100	73 73 82 85	16 16 13 12	6 7 8 5	10 9 5 7	11 11 5 3	6 7 3 3	5 4 2 -
<u>Age of persons</u> Estimated number of persons	11,527	9,196	1,563	741	822	768	525	243
Percent distributions All persons	100	80	13	6	7	7	5	2
Under 15 years 15-44 years 45-64 years 65 yrs. and older	100 100 100 100	85 84 76 73	12 12 14 17	5 5 6 12	7 7 8 5	3 4 10 10	2 3 7 6	1 1 3 4

Field Status of Dwelling Units, Families and Persons in Estimated Sample; by Size of Family and Age of Persons

1/ Interviewed in November 1960 through April 1961.
2/ The results of the second phase of interviewing on a one-third sample of cases not completed in the first phase, weighted by 3, were completed in May through July 1961.

TABLE 2