

RESPONSE STYLES IN TELEPHONE AND HOUSEHOLD INTERVIEWING:
A FIELD EXPERIMENT FROM THE LOS ANGELES HEALTH SURVEY¹

L.A. Jordan, A.C. Marcus and L.G. Reeder
University of California, Los Angeles

In the past few years, there has been a great upsurge of interest in telephone surveys. Several major comparisons of household and telephone surveys have been undertaken recently -- including the 1976 Los Angeles Health Survey (LAHS) experiment which we will be reporting on today, the 1976 University of Michigan study (Groves, 1977), the University of Cincinnati study (Klecka & Tuchfarber, 1978), and the National Crime Survey and Current Medicare Survey experiments of the Census Bureau (Bushery, Cowan, & Murphy, 1978).

After the infamous 1936 Literary Digest debacle, telephone surveys were generally dis-trusted. The prediction of an Alf Landon landslide was a grave embarrassment for survey re-search, directly attributable to the socio-economic bias in the Literary Digest's sample of telephone owners. Upwards of 95% of households now have telephones, however, so socio-economic bias is no longer a serious problem with tele- phone surveys (Dillman, 1978).

There are many good reasons for wanting to use telephone surveys:

- They are usually less expensive than house- hold surveys, depending to some extent on the kind of information and sample re- quired. Costs of telephone surveys have ranged from 20% to 66% of the costs for comparable household surveys (45-65%, Groves, 1977; 50-66%, Hochstim, 1967; 20-25%, Klecka & Tuchfarber, 1978).
- It is feasible to conduct nationwide tele- phone surveys, long-distance, using a cen- trally located and supervised staff of in- terviewers. A central location leads to more control over the interviewing process, with nearly immediate detection and cor- rection of interviewer or respondent prob- lems.
- A central location also makes possible computer-assisted telephone interviewing, such as the CATI project at UCLA (Shure & Meeker, 1978).
- Time in the field can be greatly reduced with telephone methods. The NBC-Asso- ciated Press and CBS-New York Times polling organizations can deliver the marginal re- sults from short questionnaires almost overnight.

For all these reasons, telephone surveys are an attractive alternative to household interviews. It is important to determine whether the data col- lected by telephone methods are comparable to those collected by household interviews. Initial results have been very promising -- recent com- parisons of the two methods do not reveal system- atic sampling biases (Reeder, 1975, 1976).

The few differences between modes which have

been found could be a function of sampling differ- ences, interviewing method artifacts, or both. Telephone and face-to-face interaction are very different, and telephone interviewing requires a faster pace and more verbal patter to maintain the interest and attention of respondents. Pauses which can be tolerated in a face-to-face situation can seem almost interminable on the telephone.

The Los Angeles Health Survey

The LAHS is an on-going, annual or biennial survey of health behavior and health attitudes in the Los Angeles metropolitan area. In 1976, the LAHS research team conducted a household (HH: $N = 1210$) and telephone (TC: $N = 303$) survey, at the same time, and with essentially the same in- strument. In 1977, further independent household ($N = 931$) and telephone ($N = 381$) surveys were conducted.

We attempted to keep the HH and TC inter- viewing arrangements as nearly alike as possible, especially by using the same interviewers. Al- though the interviewers had many months and often years of experience conducting household surveys at UCLA, several of them did not adapt well to telephone interviewing. For the 1977 surveys, several improvements in the telephone interviewing arrangements were made, and the interviews were made from a central location with on-line moni- toring, instead of being made from the inter- viewers' homes.

In this paper, we present the results com- paring the two modes of interviewing for selected socio-demographic and attitude variables. The main focus of the paper will be on the 1976 data, and especially on the results we obtained with a battery of attitude variables.

Socio-demographic variables

First, we will look at the socio-demographic variables, as shown in Table 1. The most striking difference between the samples is the amount of missing data for income. Respondents were asked for total family income on a 15-point scale. For the HH sample, hand cards were used, but for the TC sample, the interviewer had to read the appro- priate response categories. The income question is one of the most sensitive of the socio-demo- graphic questions, and some respondents either refuse or claim not to know the answer. We found that 12% of the HH respondents and 21% or nearly twice as many TC respondents did not answer.

In the University of Michigan study, Groves (1977) also reported more missing income data from the telephone survey. He found that the missing data rate declined for subsequent tele- phone surveys, however, which he attributed to im- proved interviewing technique. Our own experience has been similar. Figure 1 shows the missing data

rates for total family income in our 1976 and 1977 surveys. In 1977, the missing data rate dropped to 15% for the TC sample, but it is still higher than the 12% rate obtained with the two household samples. It remains to be seen whether, as more experience is gained with telephone methods, the missing data rate can be reduced to the 12% obtained face-to-face. We have obtained more "Don't know" than "Refused" responses for income, and we assume that many of these "Don't know" responses are indirect refusals from reluctant respondents (Robins, 1963). A rather large proportion of the missing income data for the 1976 telephone survey was in the "Don't Know" category (14.5%), suggesting that respondents were being evasive rather than refusing the information outright.

In view of the large amounts of missing data for income, the obtained mean difference on income is difficult to interpret. The groups are not significantly different on other indicators of socioeconomic status, such as education and Duncan's socioeconomic index (SEI), so in the absence of confirming evidence of a socioeconomic difference between the samples, we conclude that the obtained income difference is probably spurious or, if real, is rather small in size.

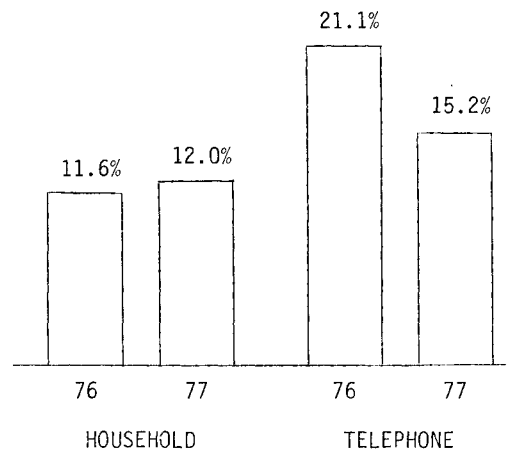


Figure 1. Missing data rates for total family income, by two survey methods, 1976-77 LAHS.

TABLE 1

Selected demographic variables - means for household (HH) and telephone comparison (TC) samples - unweighted, 1976

	HH Sample N = 1210		TC Sample N = 303		t for mean diff
	\bar{X}	S	\bar{X}	S	
Age	42.8	17.4	44.1	18.1	-1.16
% Male	0.43	0.50	0.37	0.48	1.88
% Married	0.55	0.50	0.50	0.50	1.56
N Adults	1.84	0.75	1.82	0.78	0.41
N Children	0.88	1.37	0.76	1.23	1.39
Income	8.21	3.82	7.61	3.48	2.49*
Income - % Missing	0.12	0.32	0.21	0.41	-4.58***
Mid-Education	12.46	3.57	12.25	3.36	0.93
SEI	48.80	22.50	47.83	22.74	0.67
% Anglo	0.67	0.47	0.73	0.44	2.01*†
% Hispanic	0.06	0.24	0.08	0.26	-1.27

* p < .05

*** p < .001

†Significance criterion not met for weighted analysis.

Another finding in Table 1 is that the TC sample has a significantly larger proportion of Anglos. This comparison was not significant in a weighted analysis of the same data, however, so we recommend caution in accepting the finding. The HH and TC samples were samples of households, with one respondent randomly selected from each household using the Kish (1965, p. 400) procedure. Each household had an equal chance of being in the sample, but persons in the households had unequal probabilities of selection, inversely proportional to the number of people in the household. Thus, unweighted summary statistics represent the population of households, and weighted summary statistics represent the population of persons (adults over 17 years of age). Now, despite the clear technical difference between weighted and unweighted analyses of the data, we would seldom want to make a strong inferential distinction between them. That is, for most variables of interest, we would expect an analysis of data for random respondents from randomly selected households to yield essentially the same results as an analysis of data for a simple random sample of respondents. Accordingly, we adopted the conservative practice of interpreting only the results which were robust under weighting, and were statistically significant in both weighted and unweighted analyses. In the present study, we found that the weighted and unweighted differences were almost invariably in the same direction, but that there were a few borderline results which were significant in one analysis, but not in the other.

Our final observation about Table 1 is that there are very few demographic differences between the HH and TC samples, which confirms the findings of Groves (1977) and Klecka and Tuchfarber (1978). With over 1500 respondents, we have adequate power to detect quite small differences between the samples, and our results suggest that -- apart from the greater amount of missing income data in the TC sample -- the two samples are essentially the same on all sociodemographic variables. This lack of demographic differences is a fortunate result, since telephone interviewing will probably soon be the dominant method for obtaining many kinds of survey data.

Attitude variables

The main focus of our paper today is on the HH and TC differences obtained for the attitude variables in Table 2. The first seven variables in Table 2 are 2-to-6 item content scales for measuring aspects of respondents' health beliefs (Kirscht, Becker, and Eveland, 1976; Berkanovic et al., 1978), and they are composed of 4-point Likert items (strongly agree, agree, disagree, strongly disagree). An attempt was made to balance the scales for keying in the agree-disagree direction. For example, the Acceptability scale contains the pair of items:

- Q8j "I'm very satisfied with the medical care I receive."
 Q35j "Most people receive medical care that could be better."

Persons believing that medical care is generally acceptable would tend to agree with the first

statement and disagree with the second, so that the two items can be considered balanced for agreement. Balancing is an important precaution if the items are affected by "acquiescence" or agreement response bias (Jordan, 1977). As shown in Table 2, the first four scales (from Susceptibility to Cost Concern) are balanced; the next two (Seriousness and Efficacy of Care) are slightly unbalanced, with one more disagree than agree item; and the seventh scale is unbalanced, with three agree-keyed items and only one disagree-keyed item.

The last three variables in Table 2 are response bias measures which were also obtained from the health belief items. The Agreeing bias variable was obtained by scoring all 32 items in the "agree" direction, regardless of content keying. The Evasiveness variable was computed as a count of "Don't Know" and "No Answer" responses. The Extremeness variable was computed as the difference between the number of extreme or "strong" responses and the number of middle-category responses. In raw score form, the Evasiveness and Extremeness variables were markedly skewed, so the variables were normalized by adding a constant and taking logarithms.

The main message in Table 2 is that the TC sample is higher on all three response style measures. It has a greater tendency to agree, to omit responses, and to use more extreme categories. The Extremeness effect is the strongest of the three response style effects, and is significant at well beyond the .001 level, with a mean difference on the order of one-half a standard deviation. By contrast, there are only minor mean differences for the seven content measures. There were slight differences between the HH and TC samples for the Accessibility and Motivation scales, but they were fairly small and were not robust under weighting. We did find, however, that the TC sample was more variable on all seven content measures -- significantly so for five of the seven. The larger variances are attributable to the extremeness response bias, since the extreme categories contribute more than the middle categories to variances for the content scales.

As in the case of the income question, administration of the attitude items was supplemented by a hand card for the HH sample, with the four response categories displayed for the respondent's use. The instruction read:

"I'm going to read some statements related to health and illness. I'd like you to tell me the extent to which you agree or disagree with each statement. People have many different opinions about health matters -- so there are no right or wrong answers. We are simply interested in your opinions. (HAND CARD). Please tell me if you strongly agree, agree, disagree, or strongly disagree with the following statements."

For the TC sample, the instructions were identical, but hand cards could not be used, and the interviewer instruction read:

"REPEAT RESPONSE CATEGORIES AS YOU READ ITEMS."

TABLE 2

Selected attitude variables - means and standard deviations for household (HH) and telephone comparison (TC) samples--unweighted, 1976

	Items in Scale		HH Sample N = 1210		TC Sample N = 303		t for mean diff	χ^2 for std dev diff
	Agree Keyed	Dis-Agree Keyed	\bar{X}	S	\bar{X}	S		
Susceptibility	3	3	-1.79	2.34	-2.08	2.74	1.86	12.82***
Acceptability	3	3	0.92	2.76	0.64	3.58	1.48	36.14***
Accessibility	3	3	1.97	2.34	2.31	2.84	-2.16*†	19.56***
Cost Concern	1	1	-0.45	1.42	-0.28	1.67	-1.79	13.56***
Seriousness	2	3	2.72	1.91	2.61	2.07	0.88	3.23
Efficacy of Care	1	2	3.39	1.46	3.47	1.74	-0.82	15.95***
Motivation	3	1	-4.13	1.94	-3.85	2.02	-2.23*†	0.80
Agreement Bias	-	-	12.53	3.93	13.88	4.26	-5.25***	3.25
Evasiveness (1n)	-	-	0.48	0.67	0.61	0.66	-3.03**	0.11
Extremeness (1n)	-	-	1.91	1.07	2.51	0.90	-8.99***	13.58***

*p < .05

**p < .01

***p < .001

†Significance criterion not met for weighted analysis.

Thus, although the questions were identical, much more extremeness, evasiveness and agreement response bias was obtained in the telephone survey.

These differences seem to indicate that the quality of the data obtained over the telephone was not as good as that obtained in the face-to-face situation. Higher rates of acquiescence and evasiveness in the telephone survey, for example, usually indicate that respondents are not working hard enough during the interview. As a recent NCHSR report on interviewing observed:

Answering a question accurately and completely requires the respondent to use cognitive skills in comprehending a question, recalling or organizing and processing the relevant information, and finally, in formulating an answer. These cognitive activities often require considerable effort, which the respondent must be willing to exert (Cannell, Oksenberg, & Converse, 1977, p. 14).

Exerting such effort, of course, is part of the respondent's role in survey research. Perhaps these differences pinpoint a special problem in telephone interviewing -- that of motivating people to play the respondent role.

Support for such an interpretation can be found in the literature. Consider the Michigan study for a moment -- Groves found more respondent suspicion in telephone interviews, higher refusal rates, less detailed information in response to open-ended questions, and more evasiveness to sensitive questions. Moreover, many of the specific field procedures and interviewer techniques that are traditionally used to stimulate respondent motivation (a personal style of interviewing, positive reinforcement using visual cues, introductory letters, and so on) are often more difficult to implement in telephone surveys. Telephone interviewers need to be trained as skilled technicians and motivators, since:

It is the interviewer who must make the interviewing experience and task sufficiently meaningful, sufficiently rewarding and sufficiently enjoyable to attain and maintain the necessary respondent motivation (Cannell & Kahn, 1968, p. 574).

We need research on how to train interviewers for telephone work, and on the best ways to obtain information (particularly sensitive information) over the telephone.

There are, of course, other possible

explanations for the differences that were found between the two modes of interviewing. Thus, the greater tendency to select extreme categories, acquiesce, and provide evasive answers in the telephone condition might be related to the more technical side of telephone interviewing -- to the fact that voice transmission over the telephone is imperfect, which could result in greater respondent confusion and misunderstanding; to the inability of the interviewer to use visual cues to discern when respondents need clarification; to the lack of interviewer flash cards; to the typically faster pace of telephone interviews; and so on. Indeed, it may be necessary to use quite different question formats. For example, S. Sudman (personal communication) has suggested that Likert-like information could be collected better using two questions, by first obtaining the direction (agree-disagree), and then by obtaining the strength of agreement or disagreement.

These considerations call attention to the need for further research on telephone interviewing. Clearly, far more research is needed to clarify the differences between the two modes, and to explore the dynamics that account for these differences.

FOOTNOTES

1. This investigation was supported by Grant Number 5-R18-CA-18451, "Processes in Health Behavior and Cancer Control", awarded by the National Cancer Institute, DHEW, Leo G. Reeder, Principal Investigator. Authorship of this paper is alphabetical.

REFERENCES

- Berkanovic, E., Marcus, A.C., & Jordan, L.A. The health belief model, health orientations, and the decision to seek medical care. Institute for Social Science Research, UCLA. Unpublished manuscript, 1978.
- Bushery, J.M., Cowan, C.D., & Murphy, L.R. Experiments in telephone-personal visit surveys. Paper presented at the Annual Meeting of the American Statistical Association in San Diego, August, 1978.
- Cannell, C.F. & Kahn, R.L. Interviewing. In G. Lindzey & E. Aronson (eds.) The handbook of social psychology. (2nd ed.) Vol. 2. Research methods. Reading, Mass.: Addison-Wesley, 1968.
- Cannell, C.F., Oksenberg, L., & Converse, J.M. Experiments in interviewing techniques: Field experiments in health reporting, 1971-1977. Survey Research Center, University of Michigan, 1977. DHEW Publication No. (HRA) 78-3204.
- Dillman, D.A. Mail and telephone surveys. NY: Wiley, 1978.
- Groves, R.M. An experimental comparison of National telephone and personal interview surveys. Survey Research Center, University of Michigan. Unpublished manuscript, 1977.
- Hochstim, J.R. A critical comparison of three strategies of collecting data from households. JASA, 1967, 62, 976-989.
- Jordan, L.A. The use of covariance structure analysis for the examination of acquiescence and other response style issues in faceted test data. Unpublished doctoral dissertation, The City University of New York, 1977.
- Kirscht, J.P., Becker, M.H., & Eveland, J.P. Psychological and social factors as predictors of medical behavior. Medical Care, 1976, 14, 422-431.
- Kish, L. Survey sampling. NY: Wiley, 1965.
- Klecka, W.R. & Tuchfarber, A.J., Jr. Random digit dialing: A comparison to personal surveys. Public Opinion Quarterly, 1978, 42, 105-114.
- Reeder, L.G. Recent literature concerning the use of the telephone in survey research. Institute for Social Science Research, UCLA.
- Reeder, L.G. Summary and conclusions. In L.G. Reeder (ed.) Advances in health survey research methods: Proceedings of a national invitational conference, 1975. DHEW Publication No. (HRA) 77-3154.
- Robins, L.N. The reluctant respondent. Public Opinion Quarterly, 1963, 27, 276-286.
- Shure, G.H., & Meeker, R.J. A minicomputer system for multiperson computer-assisted telephone interviewing. Behavior Research Methods & Instrumentation, 1978, 10, 196-202.