AN EXPERIMENTAL COMPARISON OF NATIONAL TELEPHONE AND PERSONAL INTERVIEW SURVEYS

Robert M. Groves, University of Michigan

This paper reports a comparison of concurrently administered telephone and personal interview surveys which attempted to collect the same information from national samples of adults. For the 1976 Spring Omnibus Survey, a personal interview sample was contacted by a staff of interviewers dispersed throughout the primary areas of the Survey Research Center's national areal probability sample; concurrently, a telephone sample was called by a group of telephone interviewers centralized in the Ann Arbor offices of SRC. The telephone interview sample was divided into two parts, both containing randomly-generated telephone numbers; one a stratified random sample of telephone households, the other a sample of telephone subscribers in the primary areas of the SRC national sample. The latter design was a feasibility test for mixed-mode surveys that would follow telephone interviews with a personal visit. The two questionnaires included identical attitudinal items on consumer finances, political affairs, relations between the races, and life satisfaction, as well as several factual items.

The discussion summarizes a large group of analyses on the data and compares the two designs on their coverage of the U.S. household population, achieved response rates, ease of obtaining interviews, demographic characteristics of respondents, differences in responses on identical questions, estimates of sampling and interviewer variance, and costs of the data collection.

1. Coverage of the U.S. Household Population by the Two Modes of Surveys

When areal probability methods are applied, errors of field listing do occur, and some members of the population are not covered by the resulting frame. For the SRC national sample of dwellings, undercoverage is estimated to include about five percent of all dwellings in coterminous United States (see Kish and Hess (1958) for a more detailed discussion of noncoverage in areal probability samples).

With random generation of telephone numbers, households in a telephone sample are identified only through their telephone numbers. If a household does not subscribe to telephone service, none of its members can be selected into the sample. The undercoverage in telephone surveys thus is concentrated in a very well-defined subpopulation. In preparation for this project we inserted a question about telephone subscription into the 1975 Fall Omnibus Survey, a national personal interview survey. We repeated that question in this project's personal interview survey and have combined the data to estimate the proportion of households that are not telephone subscribers. Table 1 shows that 7.2 percent of the households are not telephone subscribers. We emphasize that this is 7.2 percent of the respondent households; both surveys are subject to about 25 percent nonresponse. If the nonrespondent households were disproportionately nontelephone households, then our estimate of undercoverage would be low. We were sensitive to this problem and asked interviewers to record on a nonresponse form whether they were able to determine whether or not the household had a telephone. Many times the interviewers found that this was an impossible task, sometimes they made guesses about the existence of a telephone, and other times they determined this with certainty, either by observation or by asking a household member. If the nonresponse data obtained are added to those results, the percentage of households with telephone is largely unchanged.

Despite these efforts at measurement, we prefer a different data source for an estimate

Table 1

Household Telephone Ownership
by Various Household Characteristics
Combined 1975 Fall and 1976 Spring Omnibus Data

	HOUSEHOLDS WITHOUT TELEPHONE	HOUSEHOLDS WITH TELEPHONE	N
1. TOTAL SAMPLE	7.2%	92.8%	3061*
2. REGION			
Northeast	5%	95%	641
North Central	5	95	860
South	13	87	979
West	4	96	581
3. TYPE OF PRIMARY AREA			
Self-Representing Central Cities	9%	91%	234
Suburbs of Self-Representing	1	99	477
Non-Self-Representing SMSA's	6	94	1316
Non-Self-Representing Non-SMSA's	11	89	1034
4. NUMBER OF ADULTS IN HOUSEHOLD			
1 Adult in Household	12%	88%	767
2 Adults in Household	6	94	1859
3 Adults in Household	4	96	312
4 or more Adults in Household	2	98	123
5. NUMBER OF CHILDREN IN HOUSEHOLD			
0 < 18 years in Household	7%	93%	1687
1 < 18 years in Household	7	93	495
2 < 18 years in Household	7	93	465
3 < 18 years in Household	10	90	234
4 or more < 18 years in Household	10	90	176
Missing Data			4
6. RACE			
White	6%	94%	2661
Black	18	82	303
Other	12	88	86
Missing Data			11
7. 1974 FAMILY INCOME			
< $4000	20%	80%	391
$4000 - 7499	13	87	445
$7500 - 9999	10	90	283
$10000 - 14999	4	96	571
$15000 - 19999	3	97	437
$20000 - 24999	2	98	261
$25000 and over	1	99	297
Missing Data			376
1975 FALL OMNIBUS DATA ONLY			
8. HOUSING OWNERSHIP			
Home Owners	4%	96%	948
Renters	17	83	419
Neither Own nor Rent	3	97	37
Missing Data			112
9. HOUSE VALUE FOR OWNERS			
< 15000	16%	84%	161
15000 - 24999	3	97	185
25000 - 34999	2	98	167
35000 or more	0	100	318
Missing Data; Renters			685
10. MONTHLY RENT FOR RENTERS			
$50 or less	28%	72%	58
$51 - 100	26	74	119
$101 - 150	16	84	122
$151 or more	4	96	126
Missing Data; Owners			1091
1976 SPRING OMNIBUS DATA ONLY			
11. TYPE OF STRUCTURE			
Single Family House	5%	95%	1106
Other One Unit Structure	0	100	15
2-4 Total Housing Units in Structure	14	86	157
5-9 Total Housing Units in Structure	16	84	67
10 or more Total HU's in Structure	6	94	101
Trailer in Mobile Home Park	9	91	33
Trailer in Other Location	20	80	55
Missing Data			4

* 6 households of the two sample total of 3067 had missing data on the telephone ownership questions.

of the undercoverage of households by telephones. The Law Enforcement Assistance Administration National Crime Panel study interviews large samples of households each month. Response rates in the study greatly exceed those that SRC studies are able to reach. The January, 1976 panel of the survey contained about 10,000 households, 90.4 percent of which had telephones within the housing unit (Klecka, 1976). We think that this estimate of telephone coverage more accurately describes the problem faced by telephone surveys.

The ten percent noncoverage of households is double that experienced in areal probability samples, and the biasing effects of this noncoverage may be even greater because the households without telephones have very different characteristics from those with telephones. The various subtables of Table 1 show that nonphone households are disproportionately low-income, rural, rented units, likely to contain only one adult, and more likely occupied by blacks than other racial groups. The most important correlate of telephone ownership appears to be family income; telephone samples will fail to include lower income groups in their proper proportions.

The use of telephone surveys alone to infer to the entire household population is inappropriate to the extent that this undercoverage biases sample statistics. For some studies (e.g., surveys of welfare recipients) low income groups are an important portion of the population of interest, and the bias in sample statistics of a telephone survey would be large. For other purposes, when a large proportion of low income groups are not part of the study population, the bias inherent in studying only telephone households would be smaller.

2. Response Rate Analysis

Previous comparisons of personal and telephone surveys have often shown higher response rates for the telephone survey than for the personal interview portion (see Ibsen and Ballweg, 1974). Our experience has generally been the opposite. In this study the response rate for the telephone survey lies between 59% and 70% and for the personal interview survey at 74.3%. The response rate for the telephone survey is presented as a range (see Table 2) because a large group of numbers continually rang without answer when dialed. There was no way to determine whether or not these were working household numbers. The lower telephone response rate counts these as noninterview cases; the higher rate excludes them as noneligible numbers. Later work has shown that the vast majority of these numbers are nonworking, and it is likely that the true telephone response rate is close to 70 percent.[1]

Although the overall personal interview response rate exceeds that of the telephone survey, there are subsets of the population which seem to be accessed more successfully on the telephone. Traditionally, the lowest personal interview response rates are found in the largest metropolitan areas; in the twelve largest SMSA's (all primary areas of the SRC sample) the telephone interview response rate exceeds that of the personal interview (65.5 percent to 61.6 percent). Metropolitan telephone surveys may be

Table 2

Response/Nonresponse Components for Total Telephone Sample

Disposition	n	Percentages Including Ring, No Answers	Percentages Excluding Ring, No Answers
Complete Interviews	1,618	58.6%	70.4%
Partial Interviews	116	4.2	5.0
Refusal by R	203	7.4	8.8
Refusal by Other HU Member	133	4.8	5.8
Non-interview (Other)	208	7.5	9.0
R absent	21	0.8	0.9
Ring, No Answer	460	16.7	99.9%
	2,759	100.0%	

relatively more attractive than personal surveys in those areas. In addition, although the overall telephone response rate is near 70 percent, the rate for the state of Michigan, the area closest to the telephone interviewing staff, is near 80 percent. This result suggests that local telephone surveys, where the sample may have some familiarity with the research organization, may more successfully obtain interviews.

4. Characteristics of Respondents

An examination of the demographic characteristics of respondents may provide some insight into the sources of nonresponse differences between modes.[2] Differences in the distribution of respondents' race, sex, and occupation are negligible or have no clear pattern. Respondents' age and education and total family income, however, reveal consistent discrepancies between the two surveys. A larger proportion of telephone respondents are less than forty-five years of age (Table 3, 60.2 percent) than personal interview respondents (52.3 percent). Larger proportions of telephone respondents (Table 4, 76.3 percent) than personal interview respondents (70.5 percent) failed to obtain a high school diploma. Similarly a larger percentage reported total family incomes of greater than $15,000. In short there is some evidence that younger persons

Table 3

Age of Respondent by Sample Type Using Weighted Data [a]

Respondent Category	Phone	Personal (Households with phone)	Personal (Households with no phone)	Total Personal
18-24 years	16.2%	15.1%	31.5%	16.0%
25-29 years	14.0	12.0	15.2	12.2
30-34 years	10.3	9.5	11.2	9.6
35-39 years	10.3	8.2	6.7	8.1
40-44 years	9.4	7.5	7.9	7.5
45-49 years	7.9	9.5	5.6	9.2
50-54 years	7.8	7.8	5.6	7.7
55-59 years	7.2	7.7	5.6	7.5
60-64 years	6.1	6.8	6.7	6.8
65-69 years	4.7	6.0	2.2	5.7
70-74 years	3.2	4.9	0.6	4.7
75-79 years	1.4	2.6	0.6	2.5
80-84 years	0.9	1.4	0	1.3
85-89 years	0.5	0.8	0.6	0.8
90-94 years	0	0.2	0	0.2
95 or more	0	0.1	0	0.1
TOTAL	99.9%	100.1%	100.0%	99.9%
% Unweighted N	1575	1421	106	1527
MISSING DATA				
Terminated	103			
Other	56	18	3	21

[a]Data weighted by reciprocal of selection probability

Table 4

Education Summary of Respondent by Sample Type Using Weighted Data[a]

Respondent Category	Phone	Personal (Households with phone)	Personal (Households with no phone)	Total Personal
8 grades or less	8.2%	12.7%	34.1%	14.0%
8 grades or less, plus non-academic training	1.3	1.7	1.1	1.6
9 - 11 grades, no diploma	10.7	11.5	21.8	12.1
9 - 11 grades, no diploma, plus non-academic training	3.5	3.6	5.0	3.7
High School diploma	21.6	21.6	17.9	21.4
High School diploma, plus non-academic training	14.1	12.9	10.6	12.8
Some college - 1/2 year - 3 years	22.3	19.5	6.1	18.7
Junior or Community college degrees	1.6	1.9	2.2	1.9
BA level degrees	11.9	10.7	0.6	10.1
Advanced degree including LLB	4.9	3.9	0.6	3.7
Don't Know	0	0	0	0
TOTAL				
%	100.1%	100.0%	100.0%	100.0%
Unweighted N	1607	1431	107	1538
MISSING DATA				
Terminated	103			
Other	24	8	2	10

[a]Data weighted by reciprocal of selection probability

5. Response Differences Between Modes

A comparison of response distributions from the two modes in this project can suggest topic areas or question types that may be better measured in one mode than the other. We cannot estimate the pure effect of administration mode because two different interviewing staffs conducted the surveys, because each survey is subject to its own nonresponse problems, and each survey covers different portions of the U.S. household population. The latter complication can be alleviated by comparing the telephone respondents with those personal interview respondents in telephone households. Even with this control, however, we can only contrast two bundles of methodologies, each with its own collection of errors and effects of administrative organization.

Over two hundred different measures common to both modes were obtained; only a few statistically significant differences between modes were obtained. Some differences that are visible suggest weaknesses in the telephone survey data. Missing data due to failure of the respondent to answer or of the interviewer to ask the question were found to be somewhat higher on the telephone than in face-to-face interaction. On later SRC telephone surveys asking the same questions, we found that the missing data rate on the telephone survey declined over time to very near that of the personal interview survey. The result supports the hypothesis that a telephone interviewing staff can improve with experience.

Another weakness in the telephone survey data appears on open-ended items where fewer respondents offer several different thoughts in response (see Groves, 1976). One question was inserted in both questionnaires specifically to investigate this problem. A list of important problems facing the country was requested, and the probing to be used by interviewers was written into the instrument. About eleven percent fewer telephone respondents than personal interview respondents supplied three or more problems. In multivariate analysis of this measure, younger, more affluent respondents and those judged more interested in the interview were found to exhibit the largest differences between mode. When it was noted that the telephone interviews generally were faster paced, the conjecture was made that these groups, who often supply full and detailed answers, might more quickly adjust their behavior to the faster pace.

Another indicator of potential problems in the telephone survey data arose from attitudinal measures gauging the respondent's reaction to the interview. Fewer telephone respondents (39.4 percent) preferred that mode of answering questions (relative to face-to-face or self-administered questionnaires) than did personal respondents prefer the face-to-face mode (78.4 percent). Proportionately more telephone respondents noted that they felt "uneasy" about discussing some topics, especially their financial status and political attitudes. The telephone interviewers observed more suspicion and questions about the legitimacy of the study than did personal interviewers.

Other differences that exist do not suggest weaknesses in one of the modes but rather the effects of varying constraints in the two modes. Questions utilizing response cards in the face-to-face interviews were adapted to the telephone in a variety of ways. We found that the differences between modes on these questions seem to be sensitive to how many points on the scales are labelled, whether the scale is numerically-based (e.g., income, years of education). Method effects also depend on whether the telephone interviewer presents the entire scale or first its major categories (e.g., agree, disagree) followed by more specific categories (e.g., strongly, weakly disagree).

We found little evidence of different responses to items with socially desirable answers (see Hochstim, 1967; Colombotos, 1965). Although there is some evidence of greater respondent optimism on the telephone for consumer sentiment items and life satisfaction items, later surveys suggest that this was not a reliable result. Consistent with past results (Rogers, 1976), negligible differences between modes were found on reports of voting behavior.

Although we found few differences between mode on the total sample, many analyses on such data use statistics calculated on subclasses. Using age, education, income, and race groups, we searched for subsets of the population that might reveal differential effects of mode. This was largely unsuccessful; the differences were usually within sampling error and somewhat unstable across measures.

6. Calculation of Sampling Errors

In all three of the sample designs used in this project, sampling variance arises from two different sources, differences among persons that happen to be selected on different draws of the sample and differences of sample size achieved in different draws. In addition, random-digit dialed samples experience sample size variation because they search for a subset of all ten-digit telephone numbers. There is no control on what proportion of sample telephone numbers are working

Table 5

Sampling Error Calculations for Stratified Phone, Clustered Phone, and Total Personal Interview Samples

Variable Description	Mean Value or Proportion of Adults			N			Square Root of Design Effect				Coefficient of Variation		
	Stratified	Clustered	Personal	Stratified	Clustered	Personal	Stratified	Clustered	Personal	Personal Reduced	Stratified	Clustered	Personal
Reporting they live in a rural area	.19	.21	.34	790	829	1548	1.10	1.20	.93	.96	.0376	.0354	.0259
Reporting they live in or near a city of 50,000 or more	.39	.38	.66	720	759	1548	1.07	1.13	.93	.96	.0394	.0351	.0259
Reporting that they itemized deductions on 1975 tax return	.53	.53	.47	750	789	1486	1.07	1.11	.95	.97	.0387	.0358	.0286
Feeling Satisfied to Completely Satisfied about life as a whole	.84	.83	.83	402	401	723	1.02	1.27	1.00	1.00	.0533	.0660	.0341
Reporting total family income less than $7,500	.19	.20	.26	662	703	1348	1.00	.80	1.05	1.03	.0414	.0378	.0334
Feeling Mostly Satisfied, Pleased, or Delighted about life as a whole	.80	.79	.86	393	435	811	1.05	1.09	1.09	1.04	.0538	.0505	.0325
Feeling better off financially now than 1 year ago	.38	.38	.36	837	865	1531	1.07	1.23	1.09	1.05	.0366	.0456	.0252
Reporting that they planned to vote in 1976 Presidential Election	.85	.86	.78	780	796	1548	1.04	1.28	1.11	1.06	.0379	.0390	.0259
Reporting that they were not presently working	.37	.35	.42	799	838	1547	1.07	.97	1.19	1.10	.0374	.0369	.0258
Feeling saving money more important now than usual	.64	.59	.68	800	837	1494	1.05	1.46	1.20	1.11	.0374	.0451	.0264
Reporting that they voted in 1972 Presidential Election	.65	.70	.62	787	814	1507	1.12	1.32	1.18	1.10	.0377	.0389	.0269
Feeling "Very Happy" these days	.34	.30	.30	794	821	1521	1.07	1.41	1.21	1.12	.0376	.0412	.0257
Not obtaining at least a high school diploma	.22	.25	.31	779	827	1538	1.07	1.37	1.21	1.12	.0379	.0345	.0262
Mean feeling thermometer rating for Gerald Ford	52.90	52.96	54.29	734	769	1485	1.02	1.00	1.22	1.13	.0391	.0428	.0271
Who are 18-29 years old	.31	.30	.28	769	806	1527	1.12	1.06	1.26	1.15	.0382	.0343	.0253
Thinking of themselves as a Democrat	.49	.53	.53	759	794	1516	1.08	1.25	1.28	1.16	.0386	.0357	.0260
Feeling Whites have right to keep Blacks out of their neighborhood	.06	.07	.10	784	812	1525	1.06	1.41	1.34	1.19	.0378	.0376	.0262
Mean feeling thermometer rating for Jimmy Carter	54.57	55.26	57.53	616	630	1290	1.05	1.18	1.46	1.31	.0430	.0463	.0309
Mean number of telephones in home	1.89	1.92	1.73	800	838	1546	.78	1.00	1.54	1.31	.0374	.0375	.0258
Mean number of problems facing the country	3.99	4.02	4.28	775	826	1535	1.06	1.22	1.61	1.35	.0380	.0397	.0261
Who are nonwhite	.13	.13	.14	782	818	1545	1.06	.99	1.62	1.36	.0378	.0342	.0260
Feeling Cockroaches are not a problem in their home	.73	.76	.75	798	836	1546	1.07	1.37	1.74	1.44	.0374	.0372	.0258
Mean over 22 variables							1.05	1.19	1.24	1.12			

household numbers. In this project about 22 percent of all sample numbers were household subscriptions, but other samples could have by chance experienced a higher or lower proportion of eligible numbers. This source of variation in sample size is present in both telephone samples. Finally, the sample size of the clustered telephone design varies for one additional reason. Some telephone exchanges serve both households within and outside a primary area of the SRC national sample. Telephone numbers selected from these exchanges were screened, and in total we found that about seventy percent of them serve households within the primary area. Unfortunately, there is no control on this proportion and it could vary over different sample draws creating different totals of eligible household numbers generated.

Table 5 presents sampling errors for the statistics calculated on the total sample.[3] All statistics are proportions of the total sample except for those that are labelled as mean values. We present four separate pieces of information for each sample type: the mean value or proportion of adults having such a characteristic, the unweighted number of observations, the square root of the design effect, and the coefficient of variation of cluster size. All means and proportions are calculated using the selection weights arising from variation in number of eligible respondents in the sample household. The design effect, deff, is presented as a measure of the relative precision of the means and proportions. The square root of deff (called deft) is the ratio of the two standard errors. Since many packaged computer analysis programs produce estimates of variances or standard errors based on the assumption of simple random sampling, deff's or deft's can be used as multiplicative adjustments to these values to calculate the appropriate sampling error or to adjust confidence intervals to account for the complexities of the sample design. By comparing the variance of the design to that of a simple random sample of the same size, deff's also adjust for differences in the number of interviews in each sample. For the stratified random telephone sample, a deft greater than 1.0 or increased variance relative to a simple random sample of the working household numbers arises from the lack of control of sample size, and for the clustered telephone sample, from both lack of control on sample size and clustering effects. For the personal interview sample, deft's greater than 1.0 arise from the effects of clustering.[4] We expect the deft's for the stratified random telephone sample to be lower than those of the clustered telephone sample for the same statistic.

The final section of Table 5 presents coefficients of variation for the cluster size in the three different designs. All three samples have coefficients of variation safely below the level threatening the ratio mean variance approximation, (they range from .02 to .05), but the figures do provide evidence for the increased variability in size within the telephone samples (about a 40% increase in the coefficient of variation). This

reflects the variation in proportion of working numbers across the central office codes sampled.

The deft's in Table 5 are arranged by their value within the personal interview sample from lowest to highest.[5] Using the reduced personal sample, the range is .97 to 1.44 with an average over the twenty variables of 1.16. For the clustered telephone sample the order of estimates by the deft values is somewhat different, but the range of values is .80 to 1.46, with a mean deft of 1.19. The stratified telephone sample in general has the lowest design effects, a range from .78 to 1.12, and a mean deft over the twenty proportions of 1.05.

We are reminded by this exercise that although the clustered telephone sample is probably subject to less control over sample size than the personal interview sample in the same primary areas, telephone sampling within primary areas selects elements directly, all over the area, while the personal interview sample further clusters the sample into secondary units. The added clustering within primary areas in the personal interview sample may produce higher design effects than an element sample spread over the entire area. Thus, the effects of lack of control over sample size in the telephone sample may be nearly balanced by the secondary clustering effects in the personal sample.

Comparing the stratified and clustered telephone designs, we observe an average 14 percent increase in the standard error for the clustered sample. That reduced precision added to the forty to fifty percent increase in sample numbers required in the clustered sample makes the clustered design more attractive only for studies planning later personal interviews in the same households or studies of change from estimates obtained in other studies in the SRC primary areas.[6]

7. Interviewer Effects Within the Telephone Survey

One source of nonsampling error can be linked to the interviewers. Past research has demonstrated that individual interviewers may, because of different styles of asking questions, personality differences, or interactions of respondent and interviewer characteristics, produce different responses from the same respondents (e.g., Hanson and Marks, 1958; Dohrenwend et al, 1968). Following the approach of Hansen, Hurwitz, and Madow (1953), we characterize the effect of interviewer differences on the variance of a sample mean or proportion as a design effect:

$$\text{Deff}_{int} = 1 + \rho_{int} \quad (b_{int} - 1)$$

where ρ_{int} is a measure of within-interviewer homogeneity, reflecting the extent to which answers of an interviewer's respondents resemble one another, and where b_{int} is the average number of interviews taken by an interviewer.[7] This design effect measures the change in the variance of sample estimates due to the fact that clusters of respondents were interviewed by the same person instead of by different people. If there are interviewer effects on responses, respondents of the same interviewer will tend to give distinctive answers, ρ will be positive and the $\underline{\text{deff}}_{int}$ will be greater than one.

In order to calculate $\underline{\text{deff}}_{int}$, the interviewers must be selected at random from among those available, and be assigned sample elements at random to eliminate any covariation of interviewer attributes with respondent attributes. Randomized selection of interviewers from among those judged eligible did not occur; indeed the selection process attempted to achieve a uniformly high interviewer quality, and homogeneity, rather than heterogeneity, across interviewers would be the expected result of the personnel decisions. The effect of this departure would presumably decrease interviewer variance and our analysis will probably err on that side. Conversely, in terms of inference to later project experiences, the personnel decisions will probably be repeated, and this project's results are useful guides to later results. The second requirement for estimating $\underline{\text{deff}}_{int}$, the randomization of assignment of sample elements to interviewers, was painstakingly implemented in the project. As part of the sampling process, equal-sized subgroups of the sample were randomly assigned to interviewers so that, in essence, each interviewer was responsible for a small national sample. Since the telephone interviewers worked specific hours within each day, however, they could not make calls on numbers at all hours, and periodically sample numbers were randomly reassigned manually to interviewers that worked different shifts. What results from the process is a randomization within interviewer shift. Because of this, the $\underline{\text{deff}}_{int}$ measured will also contain differences between the types of interviewers that work different shifts and respondents reached during different shifts. We suspect that respondent differences across shifts are largest between those reached on weekday mornings and afternoons on one hand, and those reached on weekday evenings and weekends. An examination of the personnel on each shift shows that about two-thirds of the interviewers work in both of these groups, and we have collapsed over shifts in the analysis that follows.

Values of ρ^*_{int} were calculated for the twenty-four estimates; their values range from -.01 to .07.[8] The highest ρ^*_{int} (.071) corresponds to the number of problems facing the country mentioned by respondents. This number is probably affected by the quality of probing used by the interviewer. We noted earlier that respondent behavior regarding this question seems to differ by mode of interview. Other estimates subject to high interviewer variance are the proportion feeling that it is more important than usual to add their savings (an open ended attitudinal measure, $\rho^*_{int} = .045$) the proportion who report that they are not currently working (a sensitive subject to some respondents, $\rho^*_{int} = .038$), the percentage of respondents who did not reveal their total family income (either directly or by responding to the trichotomous categorization of income, $\rho^*_{int} = .027$). Two estimates arise from the same questions as two of the above but have much lower interviewer effects. The proportion of respondents whose total family income was less than $7,500 has a small positive ρ^*_{int} (.003), and the proportion of respondents who fail to mention any problem facing the country has a small negative ρ^*_{int} (-.001). The discrepancies in interviewer effects between the two estimates related to total family income could

support the hypothesis that reluctance to provide income to the interviewer may result from interviewer inflection or hesitation in asking the question (a variable over interviewers); once committed to giving an income figure, the proportion who reveal a low income (less than $7,500) is rather stable over interviewers. The questions asking for a listing of the most important problems facing the country should have a different pattern; we would expect relatively large interviewer effects both for the mean number of problems mentioned and the proportion of respondents who cannot identify any problems. The former is highly variable over interviewers (ρ^*_{int} = .071), but the rate of "don't know" on the item is fairly stable (ρ^*_{int} = .001). It may be the case that initial delivery style of the question has little effect on the probability of a respondent mentioning at least one important problem. In contrast although the probing was specified in the questionnaire, the number of problems mentioned seems to be much more dependent on interviewer style.

These results inform us about interviewer effects in this telephone survey, but we cannot present a comparable analysis for the personal interview survey. Despite this, a comparison of telephone interviewer effects with those of previous personal interview surveys may give some insight into the relative magnitudes of interviewer variance in the two modes. To do this we utilize three published studies: 1) Hanson and Marks' (1958) analysis of enumerator variance in 21 counties of Ohio and Michigan during the 1950 U.S. Population Census, 2) Kish's (1962) study using two surveys of factory workers, and 3) Freeman and Butler's (1976) study using a survey of urban housewives.

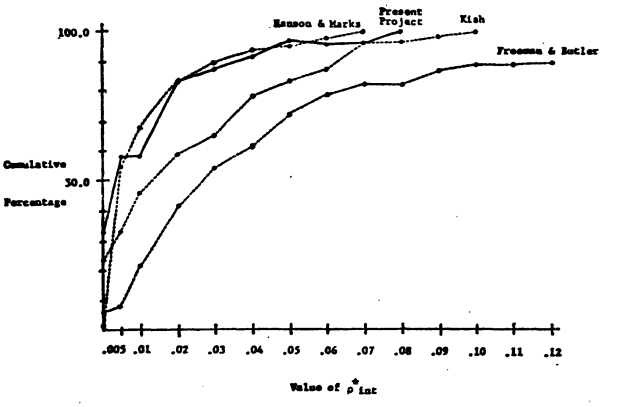
The census study includes purely demographic measures, some sensitive like income, others without any threat to the respondent, like sex; still others measure missing data on schedules returned by the interviewers. The Kish study yielding the largest range of ρ's, asked attitudinal questions about union activities and job satisfaction, in addition to some purely demographic measures. The Freeman and Butler study calculated ρ's on all categories of seventeen different variables, some of them attitudinal variables related to the general topic of mental retardation among children, others are reports of their actions toward their own children, or reports on personal behavior of other kinds.

Interviewers in the Census study were those paid as enumerators in that Census, Kish's studies used professional male interviewers employed by the Survey Research Center, and the Freeman and Butler study used school teachers none of whom had interviewing experience, but who participated in a "three-credit-hour university course in interviewing conducted by the project and field directors" (p. 84).

Figure 1 presents cumulative percentages for values of ρ's and ρ^*_{int}'s for the four different studies. The results of the telephone study are plotted with the solid black line. The highest ρ's are those found by Freeman and Butler's study of housewives. The Census study has the smallest ρ's, although our telephone survey produces the largest proportion of ρ^*_{int}'s less than zero. The Freeman

Figure 1

Cumulative Percentage Distribution for ρ^*_{int} Values Measuring Interviewer Effects for Three Personal Interview Surveys* and this Telephone Survey



* Hanson and Marks (1958) and Kish (1962), Study 1, distribution taken from Table 2, p. 97 of Kish. Freeman and Butler (1976) distribution taken from their Table 1, pp. 86-87.

and Butler study exhibits interviewer effects much higher than any of the other studies, and the use of new, nonprofessional interviewers may be associated with that result. Even ignoring that result, however, it appears that the interviewer variance experienced in the telephone survey is often lower than that in the personal interview surveys included in Figure 1.

Although the inference from Figure 1 is complicated by variation in type of measures, interviewers and populations, it suggests that telephone interviewer effects measured by ρ^*_{int}'s may be somewhat smaller than those in personal interviews. The important lesson of Figure 1, however, requires additional information. As we noted earlier, the effect on the variance of sample estimates corresponding to interviewer differences can be characterized as:

$$deff_{int} = 1 + \rho^*_{int}(b_{int} - 1)$$

where b_{int} is the average number of interviews taken by an interviewer. We have presented ρ^*_{int}'s in order to control differences in the workload of interviewers across the different studies. This is a proper approach when comparing the magnitude of interviewer variation in the two modes, but it ignores possible administrative differences in the modes. In the telephone survey interviewers each completed an average of forty-four interviews; the corresponding number in the personal interview survey is eleven. With a ρ^*_{int} of .04, which is likely in both surveys for an open-ended or sensitive item, the $deff_{int}$ for the telephone survey is 2.72; for the personal, 1.40. Simply because the telephone interviewers each take more interviews, the loss of precision arising from interviewer effects is larger. Indeed, the interviewer differences measured by ρ^*_{int} have to be less than one quarter their size in the personal interview survey for the design effects due to interviewer differences to be the same. The results in Figure 2 suggest that this will not always be the case. This illustrates that interviewer effects within centralized telephone interviewing facilities may be a larger threat to survey precision than in dispersed personal

interviewing situations. The very fact that all telephone interviewers work in the same location, and that there are relatively few of them, however, facilitates the study of methods to reduce interviewer variance in ways not possible in personal interviewer studies.

The data on sampling and interviewer variance should be combined to provide estimates of change in standard errors of the telephone survey as we administered it from one yielding the same estimates from a simple random sample interviewed singly by different interviewers.[9] The columns in Table 5 listing the square roots of design effects and ρ's for sampling and interviewer differences can be used to provide an overall effect. Table 6 presents an ordering of the overall deft's for the twenty-two estimates common to the sampling error and interviewer variance analysis separately for the stratified and the clustered telephone samples. The deft's range from .74 to 1.57 in the stratified sample and .83 to 1.75 in the clustered sample. This implies a 60-75 percent increase in the width of confidence intervals for some sample statistics. For those variables sensitive both to clustering and to interviewer effects (e.g., attitudes about the need for saving money), the total design effect is rather large (deft = 1.75), but in some cases the interviewer effects actually decrease the overall design effect from that due to sampling alone. This overall design effect may be a more proper inflation factor for simple random sample standard errors that are produced by most packaged computer programs.

Table 6

Estimated Overall Design Effect Including Sampling and Interviewer Variances for Twenty-two Measures by Telephone Sample Type

Variable Description	Square Root of Design Effects				Square Root of Overall Design Effect	
	Stratified Sample		Clustered Sample			
	Sampling	Interviewer*	Sampling	Interviewer*	Stratified Sample	Clustered Sample
Reporting total family income less than $7,500	1.00	1.02	.80	1.03	1.02	.83
Mean number of telephones in home	.78	.97	1.00	.97	.74	.96
Who are 18-29 years old	1.12	.92	1.06	.92	1.06	.98
Who are nonwhite	1.06	1.04	.99	1.04	1.09	1.03
Mean thermometer rating for Gerald Ford	1.02	1.10	1.00	1.11	1.11	1.10
Mean thermometer rating for Jimmy Carter	1.05	.91	1.18	.91	.97	1.11
Reporting they live in or near a city of 50,000 or more	1.07	.98	1.13	.98	1.05	1.12
Reporting that they itemized deductions on 1975 tax return	1.07	1.02	1.11	1.02	1.09	1.13
Feeling Mostly Satisfied, Pleased, or Delighted about life as a whole	1.05	1.04	1.09	1.05	1.09	1.14
Feeling better off financially now than one year ago	1.07	.99	1.23	.99	1.06	1.22
Feeling Satisfied to Completely Satisfied about life as a whole	1.02	.98	1.27	.98	1.00	1.25
Reporting that they planned to vote in 1976 Presidential Election	1.04	1.04	1.28	1.05	1.08	1.31
Reporting that they were not presently working	1.07	1.32	.97	1.33	1.37	1.32
Reporting they live in a rural area	1.10	1.13	1.20	1.13	1.22	1.32
Thinking of themselves as a Democrat	1.08	1.09	1.25	1.10	1.16	1.33
Not obtaining at least a high school diploma	1.07	.97	1.37	.96	1.04	1.34
Feeling cockroaches are not a problem in their home	1.07	1.01	1.37	1.01	1.08	1.37
Feeling "Very Happy" these days	1.07	1.01	1.41	1.01	1.08	1.41
Reporting that they voted in 1972 Presidential Election	1.12	1.13	1.32	1.13	1.23	1.42
Feeling Whites have a right to keep Blacks out of their neighborhood	1.06	1.17	1.41	1.18	1.22	1.54
Mean number of problems facing the country	1.06	1.53	1.22	1.56	1.57	1.71
Feeling saving money more important now than usual	1.05	1.37	1.46	1.39	1.41	1.75

* Those deft's were estimated by using the ρ^*_{int} values presented in Table 5.4 and the number of valid responses in Table 5.2.

8. Sampling and Data Collection Costs for the Surveys

The small literature that does exist regarding telephone surveys frequently contains references to costs associated with the method. Coombs and Freedman (1964) estimate that using the telephone wherever possible in a reinterview of respondents "resulted in savings of approximately 60 percent." The field cost per five-minute telephone interview of the national sample of Kegeles et al (1969) was about six dollars which they labeled "only a fraction of what a personal interview would cost." Hochstim (1967) incurred telephone interviewing costs which were fifty to seventy percent of those for the same interview completed in person. Tuchfarber and Klecka (1976) estimate personal interview costs at five times the costs for a comparable RDD survey of Cincinnati households.

Before we describe our methods of cost analysis, we should outline several dangers of inference from the costs of any one project. Each survey has unique characteristics which affect its total costs: the nature of the population studied, the size of the sample, the length and complexity of the questionnaire, and the number of interviewers employed. This project itself has some characteristics which may or may not be duplicated in future studies of either mode. This was the first telephone survey with randomly generated sample numbers ever conducted by the Survey Research Center; new methods, however pretested, inevitably bring with them difficulties of administration. Since this project we have completed other such telephone surveys and are enjoying greater efficiency in some areas than we did earlier. Also, because of the methodological nature of this telephone survey, the research staff had a larger involvement in the interviewing process than in later telephone surveys, and its participation no doubt reduced the activities of the field office personnel. Other qualities of the two different surveys, while each typical of the particular mode, may complicate the comparison of costs between modes. For example, the average personal interview lasted about fifty minutes, the telephone, only thirty minutes.[10] All these complications limit the utility of our data to other researchers for judging costs of either survey mode. We have chosen not to adjust costs in the two surveys in an attempt to reduce differences; rather we will present costs actually incurred by the two modes.

Table 7 summarizes the direct costs for sampling and field activities on the two studies. The table is broken into ten categories, representing major divisions of work. Costs for all items, person hours for salary items, and unit counts for non-salary items are listed for the components of each category.

Table 7

Direct Costs for Components of Sampling and Data Collection Activities on the Telephone and Personal Interview Surveys

	Telephone Survey		Personal Interview Survey	
	Hours or Other Units	Costs	Hours or Other Units	Costs
I. Sampling				
Administrative Salaries	86.0*	$ 505.27*	362.0	$3,305.03
Clerical/Typing Salaries	0	0	186.0	676.12
Chunking and Listing		0		4,366.00
Data Processing		450.00		0
Category Total	86.0 hours	$ 955.27	548.0 hours	$8,547.15
Percentage of Total	1.6%	2.5%	4.1%	10.1%
II. Pretest				
Ann Arbor Field Office Salaries	34.0	$ 224.59	32.0*	$ 200.08*
Clerical/Typing Salaries	4.0	14.73	17.0	88.65
Supervisors Salary	50.5	188.67	25.8	153.30
Interviewers Salary	76.4	226.06	125.7	444.13
Travel	0	0	(666mi)	93.32
Duplicating	(1,688p)	69.40	(7,370p)	119.62
Postage		0		14.00
Category Total	164.9 hours	$ 723.45	200.5 hours	$1,113.10
Percentage of Total	3.0%	1.9%	1.5%	1.3%
III. Training and Prestudy Work				
Interviewing Supervisors Salaries	37.1	$ 155.57	667.0*	$3,929.14*
Interviewers Salaries	314.6	936.58	1,621.3*	$5,285.51*
New Interviewer Training		660.00		0
Duplicating	(324p)	26.93	(6,725p)	85.10
Supplies		16.78		223.86
Coding Staff Salaries	11.0	94.46	0	0
Coding Evaluation of Questionnaires	40.0	178.00	0	0
Category Total	402.7 hours	$2,066.34	2,288.3 hours	$9,523.61
Percentage of Total	7.4%	5.4%	16.9%	11.2%
IV. Materials				
Questionnaire	(100,800p)	$ 802.25	(224,000p)	$1,466.51
Other Data Collection Instruments		278.70	(24,840p)	704.65
Data Collection Related Materials and Reporting Forms		274.68	(30,900p)	1,211.24
General Supplies		19.33		277.75
Category Total		$1,374.96		$3,660.15
Percentage of Total		3.6%		4.3%
V. Ann Arbor Field Office				
Administrative Salaries	156.0	$1,222.48	324.0	$2,508.29
Clerical/Typing Salaries	55.0	172.26	392.4	1,651.13
Category Total	211.0 hours	$1,394.74	716.4 hours	$4,159.42
Percentage of Total	3.9%	3.7%	5.3%	4.9%
VI. Field Salaries				
Supervisor Salaries	648.0	$2,303.64	988.5*	$4,956.88*
Interviewer Salaries	3,442.0	10,181.05	8,389.8*	27,321.04*
Foreign Interviewers Salaries	(6int)	60.0	0	0
Category Total	4,090.0 hours	$12,544.69	9,378.3 hours	$32,277.92
Percentage of Total	75.5%	33.1%	69.4%	38.0%
VII. Field Staff Travel				
Supervisor Travel		0		$5,620.35
Interviewer Travel				
Personal Auto Mileage	0	0	(74,405mi)	10,416.72
Other		0		778.04
Category Total		0		$16,815.11
Percentage of Total		0		19.8%
VIII. Communications				
Postage		0		$3,491.03
Telephone				
For Data Collection		$15,793.60		0
For Other Communications		0		1,756.45
Supplies For Mailing		0		732.83
Category Total		$15,793.60		$5,980.31
Percentage of Total		41.6%		7.0%
IX. Control Function				
Administrative Salaries	247.5	$ 830.55	8.0	$ 91.30
Clerical/Typing Salaries	0	0	188.5	766.93
Printing and Duplicating	0	0	(1,500p)	24.99
Data Processing		372.00*		0
Category Total	247.5 hours	$1,202.55	196.5 hours	$ 883.22
Percentage of Total	4.6%	3.2%	1.5%	1.0%
XI. Post Interviewing Activities				
A. Interviewer Evaluation/ Debriefing				
Supervisor Salaries	12.0	$ 44.52	0	$ 0
Interviewer Salaries	68.0	199.70	30.2*	98.62*
Ann Arbor Administrative Salaries	0	0	16.0	94.87
Ann Arbor Clerical/Typing Salaries	0	0	12.0	37.80
Duplicating	(50p)	2.50	(1,200p)	15.55
Postage		0		34.45
Category Total	80.0 hours	$ 245.22	58.2 hours	$ 281.29
Percentage of Total	1.5%	0.6%	0.4%	0.3%
B. Verification				
Ann Arbor Administrative Salaries	$ 100.1*	$ 384.78*	90.5	$ 596.44
Ann Arbor Clerical/Typing Salaries	0	0	9.0	35.13
Duplicating	0	9.30*	(1,150p)	33.90
Supply		24.79*		23.88
Postage		88.66*		104.78
Telephone		0		82.25*
Category Total	100.1 hours	$ 507.53	99.5 hours	$ 876.38
Percentage of Total	1.8%	1.3%	0.7%	1.0%
C. Report to Respondents				
Ann Arbor Administrative Salaries	3.0	$ 22.63	4.0	$ 28.70
Keypunching	34.2	256.23	32.8	246.00
Data Processing		162.32*		112.94
Printing	0	555.00*	(22,400p)	251.00
Postage		133.76*		107.62
Category Total	37.2 hours	$1,129.94	36.8 hours	$ 746.26
Percentage of Total	0.7%	3.0%	0.3%	0.9%
OVERALL TOTAL	5,419.4 hours	$37,939.29	13,522.5 hours	$84,863.92
PER INTERVIEW TOTAL	3.3 hours	$ 23.45	8.7 hours	$ 54.82

*Costs based on estimates of those personnel involved in the work, usually necessitated by different categories of work being performed by the same personnel.

Total direct sampling and field costs for the personal interview survey are $84,864. For the telephone survey, the costs total $37,939, only about 45 percent of those on the personal study. Person-hours total 13,523 on the personal mode, 5,419 on the telephone mode. For these two studies, therefore, the telephone mode is substantially less expensive, both in terms of direct costs and personnel time required. These results resemble those reported by Hochstim (1976) and Coombs and Freedman (1969).

For the two samples the per completed interview cost for sampling and field work is $55 using personal interviews, $23 using telephone interviews. This involves an average of 8.7 person hours per personal interview, and 3.3 person hours per telephone interview. Sample sizes were 1,548 for the personal interview study, 1618 for the telephone interview study.

While we assume that costs in survey areas other than sampling and field should be unaffected by differences in interviewing method, it would perhaps be helpful to consider our figures in the context of total survey costs. Analysis costs probably have the highest variation of all components, but we roughly estimate that sampling and field costs comprise about 50 to 60

percent of personal interview survey direct costs incurred before analysis. Expecting these other activities to cost the same for a telephone survey, we would estimate that 31 to 40 percent of total telephone survey costs up to analysis are attached to sampling and field work. Using these figures, we would expect that the total telephone survey costs would be 56 to 87 percent of total personal interview costs before analysis.

Table 7 identifies areas where large portions of sampling and field costs were incurred in each of the two modes and where large cost differences exist between the two modes. There are five areas that exhibit the largest differences. Sampling, prestudy, and training costs were markedly different in the two modes. Travel costs accounted for nearly 20 percent of total personal interview costs but were nonexistent on the telephone survey. Total communications costs (mainly WATS lines charges), on the other hand, formed over a third of all telephone survey charges and were three times as large as those for the personal interview survey. In both modes, interviewer and supervisor salaries accounted for about a third of all sampling and field costs.

There are two design differences in our studies which complicate cost comparisons. First, the fact that the sample sizes on the two studies are not identical makes use of a per interview cost somewhat difficult. We might wish to estimate costs for a different survey by multiplying the sample size by per interview cost, assuming constant marginal cost of a single interview across different sample sizes. It is more plausible that the cost of taking one interview decreases as the number of interviews increases. Therefore, having a larger telephone sample (N = 1,618) probably yields slightly lower per interview costs than would exist if the telephone sample size were 1,548. However, since the difference between the two sample sizes is small (70 cases) relative to total sample sizes (1,548 personal, 1,618 telephone) the effects of increased size are probably small.

A more serious design difference is the discrepancy in interview lengths on the two studies. To adjust for this difference, we counted the number of variables obtained in each mode. We enumerated non-missing data records on all variables that were the direct result of responses recorded by the interviewer. An approximate count for the personal interview is 289,400 and for the telephone, 260,500.[12] Using these estimates the per unit data costs are about $.29 for the personal and $.15 for the telephone survey (about 50 perent of the personal).

Another approach to calculating per unit costs focuses on time units instead of data units, and attempts to simulate costs of equal length interviews. Reducing the length of the personal interview questionnaire to .6 of its actual size (50 minutes to 30 minutes) would reduce costs of materials preparation (Ann Arbor field office work, typing, duplicating, printing), interviewer salaries and travel for pretest and the final interviewing, and other costs. But with a 30-minute personal interview it is doubtful that costs in any of these areas would be reduced to .6 of their present size. If we merely delete interviewer costs for twenty minutes of questioning, only about $1,700 is saved. But even if all preparation, field, and travel costs (categories II, IV, VI-VII in the table) were reduced by forty percent, the cost of the telephone interview survey would be only 64 percent of that of the personal survey.

We have presented three estimates of the relationship between telephone with personal interview costs. Using unadjusted project figures, sampling and field costs of the telephone survey were about 45 percent of those of the personal, per unit data costs were 50 percent of those in person, and per unit time probably somewhat less than 64 percent of those in the personal interviews.

9. Conclusions

This paper presented findings from an initial study comparing telephone and personal interview surveys. Some of the findings have been replicated by later studies; for example, we continue to achieve lower response rates in national telephone surveys on randomly generated sample numbers than in similar personal interview surveys. Other results may have arisen from our inexperience in administering such telephone surveys; the missing data rate on a series of questions has declined over repeated use of them. Still other results have become inapplicable because of new methodological developments; for example, new sample designs have increased the productivity of telephone interviewers and some costs have changed.

Future work can profitably concentrate on two different areas, 1) interviewer behavior that minimizes response and nonresponse errors, and 2) measurement of nonsampling errors. The identification of optimal telephone interviewer behavior has not yet been achieved; in this project we merely applied techniques found useful in personal interview surveys. However, new interviewer techniques may be desirable for telephone work. The first few moments of telephone interaction where many refusals occur, must form the analogue of a prestudy letter to respondents, the respondent's visual inspection of the interviewer and her written credentials, and all the accompanying descriptive stimuli that a personal interviewer provides a respondent. Now we are merely using trial and error methods in hopes of finding effective introductory techniques, but formal experimental work is required. We have noted that the tendencies toward fast pace in telephone interviews may be associated with more superficial responses to open-ended items. Response effects from questionning speed and interviewer prompting and probing should be formally studied.

All of these suggestions require a data collection design which permits measurement of interviewer effects. Telephone surveys with centralized interviewing staffs permit this more easily than personal interview surveys, and developments in using computer terminals to provide the survey questions to the interviewer and accept the answers of respondents imply that further measures of interviewer behavior may soon be possible. Measurability of these nonsampling errors both aids the evaluation of changes in interviewer behavior and provides the data analyst with better empirical estimates of error in the survey data.

FOOTNOTES

1. On a later survey the status of unanswered numbers was determined and about 95% of the numvers called at least twelve times were not working household numbers. Such unanswered numbers are disproportionately located in rural exchanges where lack of nonworking number recordings is most prevalent.

2. To eliminate one source of differences between modes, we compare telephone survey respondents with personal interview respondents whose households are telephone subscribers.

3. All variance calculations used the ratio mean formula; for the stratified random telephone sample, with elements as ultimate clusters; for the two clustered samples with primary areas as clusters.

4. Because the personal interview sample is larger than the clustered telephone sample, we would expect higher design effects for the personal interview sample. The increase is merely a function of the size of the clusters not of any differences in the sample design, and for that reason we created deft's for an "adjusted" personal interview sample. These figures are presented in the fourth column of the deft's section in Table 5. These were calculated using a sample size of 865, the maximum sample size for the clustered telephone sample.

5. Two estimates, those concerning the respondent's attitude about his life as a whole are measured on half samples. This artifically reduces their design effects for the two clustered samples.

6. We should note that as with most clustered samples, the effects of clustering on the precision of estimates is reduced for analysis of subclasses. For such analyses the clustered telephone sample is relatively more attractive.

7. ρ is a true intraclass correlation coefficient if b is a constant, or does not vary greatly over interviewers. The coefficient of variation of b in the telephone survey was about .09, and we view the presented ρ's as synthetic measures of intracluster homogeneity that also include some effects of varying interviewer load.

8. ρ^*_{int} values were estimated from a $deff_{int}$ using a clustered variance formula with unweighted data. Clusters in the calculations were all interviews completed by a single interviewer; no stratification of clusters was introduced into the calculations.

9. An overall design effect including both sampling design and interviewer effects is approximately

$$Deff_{overall} = Deff_{sampling} + (b_{int} - 1)\ \rho^*_{int}$$

following Hansen, Hurwitz, and Madow's model (1953, Vol. II, pp. 291-293).

10. A group of questions appearing at the end of the personal interview was dropped from the telephone survey questionnaire.

11. If broken-off interviews are included, the total telephone sample size is 1,734.

12. These figures were estimated by hand calculation of number of non-missing data cases in all question sets. Open-ended variables yield two data fields (first- and second-mentioned answers) and were counted as two variables. The figures are so close to one another chiefly because of the larger sample size in the telephone survey.

Selected References

Colombotos, J., "The Effects of Personal vs. Telephone Interviews on Socially Acceptable Responses," Public Opinion Quarterly, XXIX (Summer, 1965), 457-458.

Coombs, L., and Freedman, R., "Use of Telephone Interviews in a Longitudinal Fertility Study," Public Opinion Quarterly, XXVIII (Spring, 1964), 112-117.

Dohrenwend, B.S., Colombotos, J. and Dohrenwend, B.P., "Social Distance and Interviewer Effects," Public Opinion Quarterly, XXXII (1968), 410-422.

Freeman, J., and Butler, E.W., "Some Sources of Interviewer Variance in Surveys," Public Opinion Quarterly, XL (Spring, 1976), 79-91.

Groves, R.M., "On the Mode of Administration of a Questionnaire and Responses to Open-Ended Items," paper presented at MAPOR, 1976.

Hansen, M.H., Hurwitz, W.N., and Madow, W.G., Sample Survey Methods and Theory, II, New York: John Wiley and Sons, Inc., 1953.

Hanson, R.H., and Marks, E.S., "Influence of the Interviewer on the Accuracy of Survey Results," Journal of the American Statistical Association, LIII (1958), 635-655.

Hochstim, J.R., "A Critical Comparison of Three Strategies of Collecting Data from Households," Journal of the American Statistical Association, LXII (September, 1967), 976-989.

Ibsen, C.A., and Ballweg, J.A., "Telephone Interviews in Social Research: Some Methodological Considerations," Quality and Quantity, VII (1974), 181-192.

Kegeles, S.S.; Fink, C.F.; and Kirscht, J.P., "Interviewing a National Sample by Long Distance Telephone," Public Opinion Quarterly, XXXIII (1969), 412-419.

Kish, L., Hess, I., "On Noncoverage of Sample Dwellings," Journal of the American Statistical Association, LIII (June, 1958), 509-524.

Kish, L., "Studies of Interviewer Variance for Attitudinal Variables," Journal of the American Statistical Association, 57 (March, 1962), 92-115.

Klecka, W.R., "Potential Coverage Problems in Telephone Surveys," (Unpublished, 1976).

Rogers, T.F., "Interviews by Telephone and in Person: Quality or Responses and Field Performance," Public Opinion Quarterly, XL (Spring, 1976), 51-65.

Tuchfarber, A.J., and Klecka, W.R., Random-Digit Dialing: Lowering the Cost of Victimization Surveys, Police Foundation, 1976.

Work supported by NSF SOC (76-07519). Cost analysis done in collaboration with Barbara Thomas, of the Survey Research Center Omnibus Survey staff.