ESTIMATES OF INTERVIEWER VARIANCE IN TELEPHONE SURVEYS

Robert M. Groves, Survey Research Center, The University of Michigan
Lou J. Magilavy, Survey Research Center, The University of Michigan

Since the initial uses of interpenetration by Mahalanobis (1946) and the definition of the so-called U.S. Census Bureau model of response variance (1961), attempts to measure response variance have been limited to special experimental efforts conducted as part of error measurement analysis of censuses (U.S. Census Bureau, 1979; Fellegi, 1964) or as part of large scale personal interview surveys in a subset of primary areas (Bailey, Moore, and Bailar, 1978). These studies have usually concentrated on the enumerator or the interviewer as a source of correlation among response errors. Unfortunately, in personal interview surveys based on area probability samples large increases in costs are incurred to effect the interpenetration of the sample. Sending interviewers to randomly assigned locations usually increases travel costs and personnel time associated with traveling. In telephone surveys with centralized interviewing staffs this increase in travel costs is absent and the marginal costs of interpenetration are much lower. In such a survey environment it becomes possible to measure interviewer variance more routinely.

An earlier attempt to introduce interpenetrated designs into a telephone survey was reported in Groves and Kahn (1979), and indicated that the magnitude of synthetic intraclass correlations associated with the interviewer were somewhat lower in that survey than those found in personal interview surveys where similar data had been reported. (The estimator was similar to that used by Kish (1962)). Although the intraclass correlations appeared smaller, it is clear that the effect of the interviewer on total variance in a survey may be much higher for centralized telephone surveys than personal interview surveys because of the large number of cases completed by each interviewer. Thus in terms of a design effect due to interviewer variability:

$$Deff_{int} = [\ 1 + p_{int}(b - 1)]$$

where b is the average number of interviews taken per interviewer, and p_{int} is a correlation within groups of respondents contacted by the same interviewer. The increased workload of the telephone interviewer compared to the personal interviewer (by a factor of 2:1 up to 4:1, especially in national surveys) generally overwhelms the smaller p_{int} values in the telephone survey.

1. The Research Design

Since telephone interviewers share interviewing facilities, a large part of the administrative activities revolve around the scheduling of interviewers into well-defined work shifts. Contrary to the personal interview where interviewers are free to arrange interviewing work around the other activities of their personal lives, telephone interviewers are assigned fixed length shifts on specific days of the week. Unless all interviewers work all shifts on all days of the week, complete interpenetration of the sample cannot be attained. Instead what the administrative design permits is the randomization of interviewer assignments within shifts.

This survey was conducted with a computer-based telephone interviewing system, a design whereby interviewers used computer terminals which presented the appropriate questions to the interviewer and accepted numeric and text responses to the questions. Part of the system was designed to assign to interviewers sample telephone numbers to dial, allowing interpenetration with a minimal increase in costs.

The survey, which covered topics of health and television viewing habits, used a two-stage stratified sample of randomly generated telephone numbers following that of Waksberg (1978). A double sampling scheme was used for nonrespondents after the fifth week of the study in an attempt to reduce nonresponse bias. The total number of interviews completed with objectively selected adults within each sample household was 1054 for a response rate of about 67%, including unanswered numbers in the denominator of the response rate. Approximately thirty interviewers were employed on the survey, hired from among applicants to the job in the Ann Arbor office. Most were female (87%), below thirty years of age (77%), who had completed at least two years of college (73%). A small majority were part-time students (53%). None of the interviewers had previously interviewed for the Survey Research Center. All completed a three week training period.

2. The Model of Response Variance

The model we use is that suggested in Hansen, Hurwitz, and Bershad (1961) and elaborated by Fellegi (1964). It presents the expected value of j-th respondent, X_j as

$$X_j = X_{ij} - d_{ij}$$

an actual response given to the i-th interviewer, X_{ij}, and some random error component, a response deviation, d_{ij}. Sampling deviation is defined as $\Delta_j = X_j - \bar{X}_{..}$, the departure of the respondent's value from the sample mean ($\bar{X}_{..}$). Of interest in the measurement of response variance due to these response deviations is any facet of the design that might create correlated response deviations across respondents. One possible source of correlation in response deviations across respondents are the interviewers, who may through distinctive behavior induce similar response styles among respondents assigned to them. This leads to the following formulation for the total variance. For the sample estimate of the mean in a simple random sample, assuming independence of response and sampling deviations between interviewers,

$$\mathrm{Var}(\bar{X}..) = \frac{N-nk}{N-1}\left[\frac{\sigma_s^2}{nk}+\frac{\sigma_r^2}{nk}\right][1+(n-1)\rho_{int}+n(k-1)\rho'_{int}] + \frac{2(n-1)(N-nk)}{nk(N-n)}\,\alpha\,\sigma_s\sigma_r$$

where
- α is the correlation of sampling deviations and response deviations within interviewers
- n is the sample size for an interviewer from a population of size N
- k is the number of interviewers
- σ_s^2 is the variance of sampling deviations
- σ_r^2 is the variance of response deviations
- ρ_{int} is the correlation of response deviations by the same interviewer
- ρ'_{int} is the correlation of response deviations of different interviewers

The intraclass correlation, ρ_{int}, in that form has been estimated in a variety of ways in the past. Fellegi has shown that in a one-time survey with interpenetration that the following estimator:

$$\frac{1}{n}\left[\frac{n}{k-1}\Sigma(\bar{X}_{i.}-\bar{X}..)^2-\frac{1}{(n-1)k}\Sigma\Sigma(X_{ij}-\bar{X}_{i.})^2\right]=\frac{1}{n}[B-W]$$

has an expected value of

$$[\rho_{int}-\rho'_{int}]\sigma_r^2+\frac{2(n-1)}{N-n}\,\alpha\,\sigma_s\sigma_r$$

and thus estimates $\rho_{int}\sigma_r^2$ well if ρ'_{int}, the correlation among interviewers, and α the correlation of sampling and response deviations are relatively small. Kish used

$$\frac{[B-W]/n}{[B-W]/n+W}$$

which estimates

$$\rho^*_{int} = \frac{\rho_{int}\sigma_r^2}{\sigma_r^2+\sigma_s^2} \quad \text{if } \rho'_{int} \text{ and } \alpha \text{ are zero,}$$

the intraclass correlation reduced by the index of inconsistency, $\sigma_r^2/[\sigma_s^2+\sigma_r^2]$

We seek in this paper not to estimate the components of total variance and compare their relative contribution to total error. Instead, we attempt to understand the variability of interviewer effects across question types, respondent types, and interviewer types. This analysis is a preliminary step in the blending of research which attempts to estimate interviewer effects and that that seeks to reduce them. For these purposes we sought a measure of interviewer effects that could be compared across variables measured on different scales and across subclasses of different sizes. Among those used in the past, ρ^*_{int} used by Kish appeared to be desirable because it is unit free and thus can be used in comparison across statistics measured in very different units, and it is not dependent on the subclass size as are the correlated response variance estimates used in other studies. Throughout this paper we will refer to ρ^*_{int} as an intraclass correlation.[1]

The analyst confronts several problems when computing estimates of this form from real data:

a) There is some evidence that the magnitudes of within variance are not constant across interviewers on some statistics. This itself is a finding of substantive importance to those interested in the nature of interviewer variability. It also, however, produces some difficulties in statistical estimation of the precision of the ρ^*_{int}. We are currently investigating this phenomenon, but this paper will not address it.
b) Not all cases in the sample could be randomly assigned to interviewers. To do so with difficult respondents would threaten an increase in nonresponse bias as the cost of measuring response variance. This means that some differences across interviewer values may be due to purposive assignment of sample numbers to interviewers for specific reasons (e.g., ability to persuade respondents to cooperate, etc.).

All interviews obtained through refusal conversion efforts were eliminated from the analyses that follow. An examination of the differential interviewer variability across different shifts found no significant differences in values of intraclass correlations. Similarly, appointments that were not kept by the interviewer making them, resembled those obtained by the initial interviewer. Hence the data presented below pools across interviewer shifts and appointment statuses. These decisions were made using estimated standard errors of the ρ^*_{int} under the assumption of equal variances across interviewers.

3. Analysis of Interviewer Intraclass Correlations

3.1 Values of ρ^*_{int} on Seventy-Seven Survey Statistics.

The study variables were grouped into six categories of question type: 1) closed questions; 2) open questions; 3) number of mentions on open questions; 4) length of response on open questions; 5) interviewer

---

[1] Because the number of completed interviews, n, varies across interviewers the form of the estimator actually used was:

$$\rho^*_{int} = \frac{(B-W)/n^*}{[(B-W)/n^*]+W} \text{ where } B=\frac{\Sigma n_i(\bar{X}_{i.}-\bar{X}..)^2}{k-1} \text{ and}$$

$$n^* = \frac{(\Sigma n_i)^2-\Sigma n_i^2}{(\Sigma n_i)(k-1)}$$

observations (pertaining to problems in administering the questionnaire); and 6) interviewer observations (pertaining to characteristics of the respondents). Section 3.3 more fully defines these categories. Many of the closed questions required the respondent to answer on a four or five point scale. For example, for different types of television shows the respondents were asked if they liked the shows very much, liked them somewhat, disliked them somewhat, or disliked them very much. For this type of question the proportion and p^*_{int} value were calculated for the modal response category.

The values of p^*_{int} estimated for these 77 questions range from −.015 to .214 with a mean value of .028 and a median equal to .008. Questions involving interviewer observations, while meant to measure characteristics of the respondent or of the interview, are completed by the interviewer without any direct input from the respondent. This type of variable was found to have the highest values of p^*_{int}. If we examine the study variables excluding the interviewer observations, the remaining 58 variables have a mean p^*_{int}=.006, a median value of .004 and range from −.015 to .061. It is possible for estimates of p^*_{int} to be negative (Bailey, Moore and Bailar, 1978). In previous studies these negative values have often been presented as zeros. We have included the actual value in all our analyses to give the reader evidence of the instability of the estimates.

The small numbers of interviewers employed in a study, 30 in our case, contributes to the instability. In an attempt to summarize these unstable estimates we present mean and median p^*_{int} by type of variable in Section 3.3. These summary measures are very sensitive to our particular choice of variables. The large number of questions and the variability of the p^*_{int} values make it difficult to draw conclusions from these estimates. Some hint as to the magnitude of p^*_{int} is given by the standard F-test. Of our 77 study variables 30 were found to be significantly greater than zero ($p<.05$). This test of significance is based on assumptions of normality and equality of variances within interviewers. In many cases our data violate these assumptions and we may be underestimating the sampling variability of the p^*_{int}'s. It is useful to note that with an average workload of 38 interviews, as we had in this study, a p^*_{int} of only .014 would cause a 50% increase in the variance of the sample mean and a p^*_{int}=.027 would double the variance.

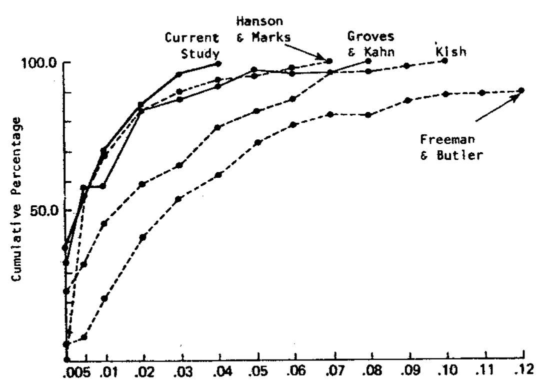
3.2 Comparison of p^*_{int} Values Across Studies

Comparison of the values of p^*_{int} found in this study with previous results gives some insight into the magnitude of the values. Figure 1 presents the cumulative percentage distribution of p^*_{int} values for two telephone surveys and three personal interview surveys. In addition to the current telephone study the comparison involves: 1) Hanson and Mark's (1958) analysis of enumerator variance in 21 counties of Ohio and Michigan as part of the U.S. Census Population; 2) Kish's (1962) study of interviewer variance in two studies of factory workers; 3) Freeman and Butler's (1976) study of interviewer effects in a personal interview survey of housewives; and 4) Groves and Kahn (1979) study of interviewer variance in a national telephone study. For the current study, values of p^*_{int} for 58 variables excluding interviewer observations are plotted. The highest p^*_{int}'s are those found by Freeman and Butler's study of housewives. Our study found p^*_{int} values lower than any of the previous studies. The earlier telephone study by Groves and Kahn also found values lower than reported by the personal interview studies of Kish or Freeman and Butler. Although inference from Figure 1 is complicated by variation in types of measures, interviewers and populations, the data suggest values of p^*_{int} found in this telephone survey are lower than those in the previous telephone survey and that telephone surveys in general may have lower p^*_{int}'s than personal interview surveys. We believe that the lower p^*_{int}'s for the telephone surveys arise from closer supervision, monitoring of performance, and exchange of interviewing techniques among interviewers.

That the p^*_{int} values are smallest for this study among all others is not an unexpected result. Interviewer behavior in this telephone study was controlled by a set of procedures built into the questionnaire. These procedures were varied experimentally, but always restricted the amount of discretion given to the interviewer regarding probing and supplementary explanation regarding the meaning of the question.

Figure 1

Cumulative Percentage Distribution of p^*_{int} Values Measuring Interviewer Effects for Three Personal Interview Surveys and Two Telephone Surveys[a]



[a] 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. Groves and Kahn (1979) results taken from Table 6.8, pp. 174-175.

3.3 Values of Intraclass Correlations By Type of Variable

One of the early findings in work on interviewer variance observed that less factual questions have higher variability across enumerators (Hanson and Marks, 1958). Fellegi's (1964) observation that measurement of whether the respondent was "French-speaking" was subject to large interviewer effects suggested a more refined explanation: the relevant dimension was not factual/nonfactual, but rather the ability of the interviewer to interfere in the recording of the data. This interference can take place through psychological influence on respondent's behavior (e.g., the sensitivity of admitting to socially undesirable characteristics), through probing behavior on questions apparently not understood by the respondent, and through differential efficiency of recording verbatim responses to open-questions. Table 1 presents summary statistics for the six different kinds of

Table 1 – Mean, Median and Range of p^*_{int} Values by Question Type[a]

Type of Question	Mean p^*_{int}	Median p^*_{int}	Minimum p^*_{int}	Maximum p^*_{int}	Number of Questions
Closed Questions	.003	.002	-.015	.023	39
Open Questions	.006	.002	-.010	.031	10
Open Questions — Number of mentions	.013	.016	-.003	.028	5
Open Questions — Length of response	.033	.028	.015	.061	4
Interviewer Observations — Pertaining to Technical Problems	.081	.047	-.004	.241	5
Interviewer Observations — Pertaining to Respondents	.096	.090	-.008	.237	14

[a] The maximum number of interviews used in calculating p^*_{int}'s was 954.

variables described in Section 3.1. The six categories have been ordered on a dimension reflecting the ability of the interviewer to affect the recorded response. Closed questions can be affected by the interviewer's inflection during the reading of the question and by any psychological influences on the respondent's behavior that vary by interviewer. These have the lowest intraclass correlations (.002 median value). The response category chosen for open-questions is additionally sensitive to the recording of the response by the interviewer; these responses have about the same median p^*_{int} value, .002. The next two types of statistics reflect characteristics of recorded responses to open-questions that are more closely dependent on interviewer actions. The number of mentions to a question may be sensitive to interviewer behavior upon delivery of the first mentions (the median p^*_{int} for these is .016); the length of the recorded response is affected by any differences in interviewer shorthand (the median p^*_{int} value is .028). By far the largest p^*_{int} values are for questions that the interviewer answers after the interview is complete. Although these interviewer observations purport to measure characteristics of the respondent or of the interview itself, Table 1 shows how dependent they are on interviewer interpretation (median p^*_{int} value is .047 and .090).

3.4 Comparison of p^*_{int} Values by Respondent Subclasses

Table 2 presents mean and median values of p^*_{int} by type of question and by sex, education and age of the respondents. The predominance of female telephone interviewers and potential differences in their interaction with female and male respondents suggested the possibility of differences in interviewer effects by sex of respondent. Table 2 shows that there is little evidence that gender of the respondent is related to interviewer effects.

Although past empirical results are mixed (Cannell, Oksenberg, and Converse, 1977; Schuman and Presser, 1977), it is reasonable to argue that poorly educated respondents might be more easily affected by the behavior

Table 2 – Values of p^*_{int} by Type of Question and Respondent Subclasses

Summary Measures of p^*_{int}	Sex		Education			Age				
	Males	Females	Less than 12yrs.	12yrs.	More than 12yrs.	18 to 24yrs.	25 to 34yrs.	35 to 49yrs.	50 to 59yrs.	60 to 100yrs.
Closed Questions										
Mean	-.002	.002	.006	.001	.002	-.004	.009	-.002	.004	.027
Median	-.004	-.004	.014	-.002	.001	.010	.011	-.003	-.002	.020
Open Questions										
Mean	.003	.005	-.000	.020	-.001	.001	-.027	.039	-.012	.020
Median	-.004	.001	.011	.016	-.012	-.018	-.020	.038	.000	.028
Open Questions Number of Mentions										
Mean	.006	.001	.014	-.001	.009	-.029	.011	.045	-.032	.045
Median	-.003	.002	.023	-.005	.012	-.031	.002	.044	-.032	.006
Open Questions Length of Response										
Mean	.019	.022	.023	.038	.030	-.030	.040	.052	.040	.038
Median	.008	.020	.028	.038	.028	.005	.048	.052	.058	.058
Interviewer Observations Pertaining to Technical Problems										
Mean	.107	.069	.050	.097	.051	.060	.087	.104	-.022	.123
Median	.080	.033	-.006	.096	.015	.074	.043	.051	-.088	.033
Interviewer Observations Pertaining to Respondents										
Mean	.085	.102	.073	.085	.121	.091	.134	.104	.064	.105
Median	.080	.088	.078	.092	.086	.080	.108	.108	.074	.084

and status of the interviewer than more highly educated respondents. Lower education groups may seek greater help from the interviewer in answering the questions or may use the inflection of the interviewer's voice as a cue for responses to questions they find difficult. Values of p^*_{int}, however, do not appear to be larger for respondents with less than 12yrs. of education.

We also expected larger p^*_{int} values for the oldest age group. There is some evidence that telephone surveys suffer greater nonresponse among older persons (Groves and Kahn, 1979), and we sought evidence that those that do respond might be more subject to influence by the interviewer because of their own suspicion about the nature of the survey, or greater tendency to fatigue during the course of the thirty minute interview on the telephone. Sudman and Bradburn (1973) have noted that older respondents exhibit greater response errors on questions requiring recall of factual material. In fact, the oldest age group tends to have larger p^*_{int} values for a majority of question types when it is compared to any age group except the 35-49 year olds. This group, 35-49yrs., has surprisingly high values of p^*_{int} for which we have no explanation.

4. Analysis of Interviewer Variation

One method of examining the nature of response differences across interviewers, an activity that goes beyond the measurement of the component of total variance due to interviewers, uses interviewers as the unit of analysis. At that level of aggregation, we can attempt to discover correlates of variability in interviewer means. It examines single correlates of the absolute value of an interviewer's deviation from the overall study mean.

It is useful to note that this analysis has two sequential steps; one searching for correlates of interviewer deviation, and the second attempting to understand the nature of the deviation. If younger interviewers achieved lower reporting on embarrassing questions, they may be introducing relative undercounts for such measures, and we would prefer the results of other interviewers. If younger interviewers were merely more highly variable about the study mean, we would have to investigate the sources of their elevated variability.

The analysis utilizes some characteristics of the interviewer (e.g., age, former work status, evaluation by supervisors) and of the interviewer's performance during the study (e.g., number of interviews completed, individual response rate). Least squares procedures are employed to model the relationships between these variables singly and the absolute value of the deviation of the interviewer's mean from the overall study mean. Because the number of interviews completed by each person varies from four to over seventy-five, a weighted least squares procedure was used. Each interviewer mean was weighted by the square root of the number of interviews completed. We used for this preliminary work four variables measured from open-questions: 1) the length of recorded response for questions concerning symptoms of health disorders, 2) the number of conditions recorded in questions about health symptoms, 3) the length of the recorded answer for questions about behaviors that affect one's health, and 4) the number of conditions mentioned in response to questions about behaviors that affect one's health. In Section 3, these statistics were seen to exhibit moderately high interviewer differences. In this analysis a maximum of twenty-nine interviewers are used. One interviewer, who completed only four interviews, was an outlier on all four variables, and was deleted from the analysis.

First we examine scatterplots of the twenty-nine interviewer means by two variables, response rate of the interviewer, and the number of interviews completed. There is a consistent negative slope for the deviation regressed on the response rate. Lower response rates are associated with higher deviation from the overall study mean. For the scatterplots of absolute deviation on the number of interviews completed, there appears to be little evidence of a relationship between the two variables. Interviewers who were very productive do not seem to deviate from the mean any more or less than do less productive interviewers (the average slopes are within $\pm$.01). The finding that deviation from the overall sample mean is a function of response rate is disturbing. One interpretation of this result is that interviewers with low response rates successfully contact a different kind of respondent than do interviewers with higher response rates. Under the argument, the departures of their respondent group means from the overall sample mean reflects not the impact of the interviewer on responses but true differences between respondent characteristics. We plan further analysis of this type to understand the nature of the relationship between deviation and response rate.

Several interviewer characteristics were measured as part of this survey. Some of them examined have little impact on deviation of interviewers from the overall study mean (e.g., whether the interviewer is a part-time student). Our analysis shows that although there are some differences suggesting that interviewers judged less adequate on typing and reading abilities achieve larger absolute deviations, few of these differences are large and none are statistically significant. Age also shows few differences. Especially noteworthy is the failure of the youngest age group (those in their early twenties) to have the highest deviations - an hypothesis of those arguing for mature telephone interviewers.

We did find that the supervisor's rating of the interviewer at the end of the study appears to be related to interviewer deviation from the overall study mean on some of the statistics. On all four variables, those interviewers rated lower tend to achieve higher absolute deviations and for the two variables concerning health behavior these differences are statistically significant at the .02 and .06 levels (using ordinary least squares formulas). This supervisory rating was made after the study, when response rates, monitoring results, and any other records about the interviewer's performance were available. The rating results are correlated to response rate (eta^2 = .53), but the ordering of response rates does not perfectly follow the order of rating categories. The ability to recognize such interviewer differences in a post-survey rating procedure suggests that some behavior of the interviewer related to their effects on responses can be observed by supervisors.

5. Summary and Conclusions

This paper presents preliminary results from an analysis of interviewer variability within centralized telephone surveys. It has replicated earlier findings that show smaller correlations of response deviations in telephone surveys than those found in personal interview surveys. The limitation of this particular work to variation among thirty interviewers makes many estimates unstable, but the potential design effect because of large interviewer workloads underscores the importance of efforts at estimating these quantities.

We are encouraged by the ability to sort measures into categories that are differentially sensitive to interviewer effects. The dimension of the interviewer's ability to influence the recorded response effectively separates measures by their sensitivity to interviewer effects. Closed questions are relatively insensitive to interviewer effects because the effects arise only through errors in delivery or differences in inflection. Just as open questions permit the respondent freedom in forming his/her answer, however, so too do they permit interviewer behavior to interfere in the process of obtaining a response (through interviewer probing) and recording the given response. Finally, interviewer observations, the statistics with the highest p_{int}'s, may be affected by the interviewer's attitude about this questionnaire in particular, about surveys in general, or about the responses desired by survey directors to the questions.

REFERENCES

Bailey, L., Moore, T.F., and Bailar B., "An Interview Variance Study for the Eight Impact Cities of the National Crime Survey's Cities Sample," Journal of the American Statistical Association, 73, 1978, 16-23.

Cannell, C., Oksenberg, L., and Converse, J., Experiments in Interviewing Techniques: Field Experiments in Health Reporting 1971-1977, National Center for Health Services Research, 1977.

Fellegi, I. P., "Response Variance and Its Estimation," Journal of the American Statistical Association 59, (December, 1964), 1016-1041

Freeman, J., and Butler, E.W., "Some sources of Interviewer Variance in Surveys,". Public Opinion Quarterly, 1976, 40:79-91

Groves, R. M., and Kahn, R.L. Surveys by Telephone, New York: Academic Press, Inc., 1979

Hansen, Morris H., Hurwitz, William N. and Bershad, Max A., "Measurement Errors in Censuses and Surveys," Bulletin of the ISI, (1961) Volume 38, No. 2, 351-374.

Hanson, Robert H. and Marks, Eli S., "Influence of the Interviewer on the Accuracy of Survey Results," Journal of the American Statistical Association, 53, (September, 1958), 635-655.

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

Mahalanobis, P.C., "Recent Experiments in Statistical Sampling in the Indian Statistical Institute," Journal of the Royal Statistical Society, 109, 326-378, 1946.

Schuman, H. and Presser, S., "Question Wording as as Independent Variable in Survey Analysis," Sociological Methods and Research, Vol.6:2, 1977, pp.151-170

Sudman, S. and Bradburn, N., "Effects of Time and Memory Factors on Response in Surveys," Journal of the American Statistical Association, 68, (December, 1973), 805-815.

U.S. Bureau of the Census, "Enumerator Variance in the 1970 Census," PHC(E)-13. (June,1979).

Waksberg, J., "Sampling Methods for Random-Digit Dialing", Journal of the American Statistical Association, 1978, 73:40-46.