THE PROCESS OF INTERVIEWER VARIABILITY: EVIDENCE FROM TELEPHONE SURVEYS

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

1. Introduction

The past literature on the measurement of nonsampling errors has two main focuses. The first is the estimation of components of total variance, sometimes total mean square error, on survey measures. From this literature has come a variety of models and estimation procedures, much of them concentrating on the measurement of effects of interviewers on the data (Fellegi, 1964; Bailey, Moore, and Bailar, 1978; Biemer, 1978). Nowhere in this treatment are discussions of the survey designs and interviewing procedures that can control the various sources of nonsampling error. The second literature investigates ways to improve the quality of survey data through questionnaire design, training of interviewers, and supervisory techniques used during the survey period. This work uses experimental variations imbedded in the survey design to test alternative procedures. Sometimes the studies have utilized administrative records in order to measure certain response and nonresponse biases associated with the different procedures (Cannell and Fowler, 1963). This literature has aggressively pursued the reduction of nonsampling errors but has generally not utilized methods that could facilitate the routine measurement of them.

This paper reports an attempt to link those two approaches to the study of nonsampling errors, specifically of those associated with effects that interviewers have on survey data. These arise, it is believed, through idiosyncratic behaviors on the interviewers' parts that create similar response errors among an interviewer's respondent group. This study used an interpenetrated design for assignment to interviewers in order to measure certain components of interviewer variance present in the data. Concurrent with the data collection by a centralized staff of telephone interviewers a monitoring procedure for the interviews was constructed so that another person listening to the interview rated the behavior of individual interviewers relative to the procedures they were trained to pursue.

The training for the interviewers specified that they were to read the questions in the instrument exactly as they appeared, with no changes in any wording. They were coached to read the instrument slowly, at a pace of about two words per second. In addition, they were given explicit instructions regarding the use of probing for incomplete answers by respondents. Finally, the interviewers were trained in voice techniques that were thought to convey a desirable professionalism. There is a strong belief among survey researchers that one source of interviewer variability can be controlled through this type of standardization. All of these rules are operationalizations of conclusions from past experimental tests of alternative interviewer procedures. These procedures formed the basis of all supervisory review of interviewers, of judgments concerning the quality of individual interviewer's work, of praise and promotions.

The design of the research project, containing both interpenetration for interviewer assignment and monitoring procedures, allows us to investigate whether the rules prescribed for interviewer behavior are related to the magnitude of interviewer variability about the overall survey statistics If we find that the amount of interviewer variability is related to specific violations of training guidelines, then we can begin to refine training procedures in order to reduce that source of error. If, on the other hand, we find that the violations of prescribed interviewer behavior result in no unusual interviewer effects, we must reevaluate our training procedures. To our knowledge this is one of the few attempts to correlate a measurable source of response variance to the definition of "correct" behavior (as judged by training instructions) on the interviewers' part.

2. Study Design

This study was designed for two purposes; first, to provide data for comparison with the Health Interview Survey face-to-face interview, and second, to explore several models of telephone interviewing. It was divided into three replicate samples introduced at the beginning of October, November and December, 1979. We employed a two-stage stratified sample of randomly generated telephone numbers following that of Waksberg (1978). A total of 8,210 interviews were completed (4,400 cooperating families) for a family level response rate of 80%, including unanswered numbers in the denominator of the response rate.

One person in each family in the household acted as a reporter on the health status of all adult family members. As part of a study of relative response errors using different respondent rules, two alternative procedures were used. In the random respondent half-sample a household listing was taken from the person who answered the telephone. One respondent from among those 17 years or older was selected using procedures similar to those of Kish (1949). In the knowledgeable adult half-sample any adult answering the telephone who judged themselves capable of answering the health questions did so for their own family. In addition, families were assigned to one of two interviewing methods ("control", modeled after the Census interviewer's behavior and "experimental", including several standardized experimental interviewing techniques) and to one of two questionnaire types (a computer-assisted telephone interviewing system and a standard paper and pencil questionnaire).

An essential condition for estimation of the parameters in the interviewer variance models discussed in Section 4 is the random assignment of telephone numbers to interviewers. In essence, this interpenetration provides each interviewer with a small national sample, thus removing the possibility of certain interviewers being consistently assigned to a particular type of respondent. In a telephone facility the physical proximity of the interviewers makes this randomization (at least within interviewer shifts) a relatively easy and inexpensive procedure. Of the 8,210 interviews obtained in the study 7,174 or about 87% were from randomly assigned phone numbers.

Thirty-three interviewers were employed on the survey, hired from among applicants to the job in the Ann Arbor office. Most were female (79%), below thirty years of age (63%), who had completed at least two years of college (97%). None of the interviewers had previously interviewed for the Survey Research Center. All completed a ten day training period.

3.1 Introduction to Monitoring

There is abundant evidence (Marquis and Cannell, 1969; Lansing, et al., 1971) that both personal and telephone interviews can be distorted as a result of an interviewer's behavior. Failure to read a question exactly as printed, inability to follow skip patterns correctly, and reading a question too fast all may contribute to errors in recorded data. Most persons working with interviewers are aware of the need for supervision to maintain the quality of an interviewer's performance. Time restrictions and lack of appropriate techniques severely limit the amount of supervision of personal interviews; the increased use of centralized telephone operations, however, has given researchers greater ability to monitor an interviewer's performance. Systematic evaluation of interviewers can be accomplished by identifying the major categories of interviewer behavior and classifying each behavior as correct or incorrect, according to the concepts and training guidelines for that particular study.

3.2 Behavior Codes

The present coding system is a revision of a more extensive system used for coding personal interviews (Cannell, Lawson and Hausser, 1977). The system is quite flexible and can be adapted to the purposes of a particular study. These codes reflect both objective and subjective measures of interviewer behavior (or lack of behavior) in five categories: 1) question asking, 2) probing, 3) defining / clarifying; 4) short feedback; and 5) long feedback. The monitor recorded whether each behavior was correct and appropriate (e.g. reads question exactly as printed reads question incorrectly - minor changes, fails to read question). For each concrete behavior the monitor was also required to evaluate the quality of the delivery. These subjective codes evaluated pace and clarity for question reading, defining, clarifying, probing, or delivering feedback.

Reliability among the monitors prior to production monitoring reached an 85% level for each of the codes in the objective categories and a 75% level for each of the codes in the subjective categories. This percentage reflects the overall agreement between the four monitors and the instructor for each of the codes.

3.3 Analysis of Monitoring Data.

This analysis seeks to determine whether interviewers differ in correct and incorrect use of techniques among themselves and whether this variance differs across questions. Although variance measures could be presented for each major category of behavior (question asking probing, feedback, etc.), we will limit the discussion to two behaviors, (a) question reading and (b) clarity and pace of question delivery. Specifically, a correctly read question is a question read exactly as written or with changes involving contractions. A question "read well" is one in which the delivery was paced at two words per second and was correctly delivered according to the following criteria:

(a) inflection. Did the interviewer's voice at the end of a phrase signify that a question, not a statement was delivered?
(b) emphasis. Should not be inadequate nor exaggerated.
(c) monotone delivery. Did the interviewer sound dull, bored, or uninterested?

Table 1. Descriptive Measures of Monitored Interviewer Behavior

Description of Statistic	n	Read Correctly: Proportion	Read Correctly: p^*_{int} [a]	Read Well (correct pace, clear speech): Proportion	Read Well (correct pace, clear speech): p^*_{int}
Two week bed days	211	.962	-.0102	.929	.0834*
Two week work loss days	214	.897	.2345*	.883	.1516*
Two week cut down days	315	.838	.0620	.895	.0858*
Two week doctor visits (from Person Section)	236	.881	.0760*	.869	.2291*
Physician visits (from Supplements)	742	.849	.0380	.889	.1526*
12 Month hospital episodes	288	.903	.0293	.910	.1297*
Two week phone calls to doctor	246	.781	.0700*	.902	.1063*
Two week dental visits	241	.979	.0822*	.938	.0676*
12 Month doctor visits	260	.719	.0448	.912	.1635*
12 Month bed days	255	.878	.-0096	.902	.2061*
Time since last doctor visit	221	.928	.0131	.919	.1046
Time since last dental visit	217	.917	.0820*	.885	.0613
Conditions					
Mean # of acute	172	.820	-.0554	.930	.1863*
Mean # of chronic	236	.674	.0136	.970	.0878*
Health status	253	.957	-.0199	.968	-.0315

[a] where p^*_{int} (see Section 4) is the intra-class correlation for monitored behavior
* significant at $p < .05$

Table 1 presents a summary of interviewer variation in question delivery for fifteen dependent variables. Even with the special emphasis given to training and continual feedback given to interviewers throughout the study, interviewers showed significant variation in their reading of a number of questions. Although some questions (dental visits, time since last dental visit) show significant variation among interviewers in the proportion read correctly it should be noted that the overall mean proportion correct is quite high (.979, .917). Also of interest are those questions which are consistently read poorly, with no variance across interviewers. The low mean proportion of correct readings for questions about phone calls to doctors, twelve month doctor visits and chronic conditions indicates that the source of error is most likely the wording of the question rather than the interviewer.

All but two of the dependent variables show a significant difference across interviewers with respect to how well the question was delivered. Coupled with the overall high proportion of questions delivered clearly and at the proper pace, we suspected that a few outlying interviewers were responsible for the consistent significant differences. Examination of the data did not support this hypothesis.

4. Response Error Model and Estimators

The response error model we employ is one that views the answer obtained from the respondent as subject to some error, a deviation from the actual value corresponding to the respondent. For example, if we are interested in the number of doctor visits for the j-th respondent, we might express the answer given by the j-th respondent to the i-th interviewer, as

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

where X_{ij} is the number of doctor visits reported to the i-th interviewer, X_j is the expected number of visits reported by the j-th respondent and d_{ij} is the response deviation committed by the j-th respondent in answer to the i-th interviewer. The expected value of the respondent reply is that "average" value obtained over all possible repetitions of the questions by all interviewers. This excludes from examination any biases that may result from procedures used by all interviewers.

Of particular interest is the pattern of response deviations (d_{ij}) that occur among the group of respondents who were interviewed by the same person. That is, we are interested in the correlation of the d_{ij} within interviewers. If there is some correlation among those response deviations, then we will view those as the effects of the interviewer on the data. With this formulation the total variance of the mean can be expressed as

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

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

As in our past work (Groves and Kahn, 1979; Groves and Magilavy, 1980) we seek to use an estimator of interviewer effects that can be compared across variables with different units of measurement. For this reason use of ratios of correlated response variance to total variance (Bailey, Moore, and Bailar, 1978) or F-statistics (Hanson and Marks, 1958; Freeman and Butler, 1976) was not attractive. Instead we use

$$\left[\frac{BMS - WMS}{n^*}\right] \Big/ \left[\left(\frac{BMS - WMS}{n^*}\right) + WMS\right]$$

where $BMS = \frac{1}{k-1}\ \Sigma n_i\ (X_{i.} - X_{..})^2 \quad WMS = \frac{1}{(n^*-1)k}\ \Sigma\Sigma(X_{ij} - X_{i.})^2$

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

which is desirable because it is unit free. The expected value of this is approximately

$$\frac{\rho_{int}\ \sigma_r^2}{\sigma_r^2 + \sigma_s^2}$$

as observed by Fellegi (1964).

In this project the same data were collected for each adult member of each family within all sample households. Generally one adult, selected in accordance with the respondent rules described in Section 2, served as the respondent for all members of the family. Consequently, in addition to the correlation of response deviations within interviewers the ρ^*_{int} values calculated for the entire sample are affected by sources of homogeneity within the family. To eliminate this component of within family homogeneity the values of ρ^*_{int} presented in Table 2 were calculated using only the "random respondents" (i.e. the randomly selected adult in each family in the random-respondent half-sample, n=1918).

5. Estimates of Interviewer Effects

Table 2 presents the mean or proportion and value of ρ^*_{int} for 15 survey estimates of health status and health-related activities. The sampling distribution of these ρ^*_{int}'s is known only under rather rigid conditions. After several studies of this type using similar designs, we expect to find some instability of the values of ρ^*_{int}'s, reflecting the small number of degrees of freedom from few interviewers. This is reflected in the several negative ρ^*_{int}'s that appear in the table. We have also noted frequent violation of the assumptions of the underlying the linear model allowing tests of hypotheses on the ρ^*_{int}'s, that is, there is some evidence that the

Table 2. Question Means and Values of p^*_{int}

Description of Statistic	Mean or Proportion (1918)	p^*_{int}
Two week bed days None	.937	-.0057
Two week work loss days None	.935	.0008
Two week cut down days None	.909	-.0006
Two week doctor visits (from Person Section) None	.849	.0092*
Physician visits (from Supplements) None	.831	.0081*
12 Month hospital episodes None	.869	-.0000
Two week phone calls to doctor None	.972	-.0020
Two week dental visits None	.936	.0040
12 Month doctor visits 2 - 4	.371	.0002
12 Month bed days None	.452	-.0070
Time since last doctor visit 2wks. - 6months	.422	.0004
Time since last dental visit 2wks. - 6months	.331	.0018
Conditions		
Mean # of acute	.173	.0004*
Mean # of chronic	.516	.0097*
Health status Excellent	.418	.0085*

magnitudes of within interviewer variance are not constant across interviewers. For heuristic guidance the starred values of the p^*_{int}'s are those that are significantly different from zero, given the assumption of equal variances.

Table 2 shows that the values of p^*_{int} for these estimates range from -.0070 to .0097 with a mean value of .0021 and a median of .0008. Although these values seem quite small, it is useful to note that such a value must be multiplied by a function of the average interviewer workload in order to estimate the inflation in total variance due to interviewer effect. In this study, we used 33 interviewers to take the 1918 random respondent interviews, thus, the average interviewer workload was 1918/33=58.12 interviews. If we estimated the inflation of variance due to the use of these 33 interviewers instead of using different interviewers for each respondent, we see that the design effect for the interviewer assignment is

$$deff_{int}=[1+p^*_{int}(58.12-1)].$$

For p^*_{int}=.0097 for the mean number of chronic conditions we calculate $deff_{int}$=1.55. In other words, we might expect a fifty-five percent increase in the variance of this estimate due to interviewer effects However, using the average value of p^*_{int} =.0021(the mean value) $deff_{int}$=1.12 a relatively small increase in variance on the average for these health variables. This finding is rather surprising given our previous work (Groves and Kahn, 1979; and Groves and Magilavy, 1980) and the work of others. However, a study of medical care utilization (Feather, 1973) in Saskatchewan, Canada reports results similar to our present findings. Although that study found significant interviewer effects for many variables, (e.g., chronic conditions, and most non-factual items related to the respondents perception of his state of health), the author concluded that individual variables measuring utilization experiences (e.g. 2 week doctor visits, 2 week bed days, etc.) are generally free of interviewer effect.

In the absence of interviewer effects the expected values for sets of respondents randomly assigned to interviewers are constant. With full interpenetration within shift like we introduced in this design, it is the case that in the absence of interviewer effects all interviewers within shift should be subject to the same expected value for their assignments. Furthermore, past analyses have shown that the effect of shift differences on interviewer means is diminished by the rotation of interviewers over shifts throughout the survey period. This permits us to compare the means obtained by different interviewers in order to observe more graphically the variation across them.

Three differences are obvious when these results are compared to those of previous work:

1) The magnitude of interviewer variation is smaller in this study than in previous studies. Past studies have shown variation in interviewer effects based on question format, with open-ended questions and attitudinal questions sometimes suffering from larger interviewer effects than factual questions. This study contained mostly factual questions about health related events that would be well-remembered by many respondents (e.g., number of hospitalizations in the last year) and only a few that require complex recall tasks (e.g " How many times have you called a doctor in the last two weeks?"). The substantive topic and format of the questions, therefore, may contribute to the overall low susceptibility of the measures to interviewer effects.

2) This study introduced an experimental interviewing procedure designed to increase the accuracy of survey results both by decreasing bias and correlated response variance. That is, it was expected that the procedures would reduce the overall tendency to underreport health events across all interviewers and standardize the interviewer behavior to reduce inter-interviewer disagreement. The half sample receiving this experimental interviewer treatment was compared to the complement half sample in which interviewers were somewhat freer to probe incomplete responses. Even in this procedure, however, interviewers were more restricted in their behavior than, for example, the Census interviewers who now administer the personal interview HIS questionnaire.

3) The form of the distribution of mean values obtained by each interviewer differs from those of past studies. The

previous telephone data yield distributions of interviewer means that contain few outliers and had relatively smooth distributions of deviations about the overall survey mean. These data, however, have many measures where one or two interviewers are extreme outliers to the distribution. Over different statistics the identity of the outliers varied. To evaluate the impact of these extreme deviations we again performed the interviewer variability analysis for five statistics with high values of p^*_{int} eliminating these outliers. For the proportion reporting no two week doctor visits one interviewer was eliminated. Two interviewers were dropped for each of the other four variables. The results of the re-analysis appear in Table 3. For each variable the measure of interviewer variability, p^*_{int} is smaller and for all variables these new values were not significantly greater than zero. In other words, one or two interviewers are responsible for most of the measured variability.

Table 3: Values of p^*_{int} for Five Selected Statistics Before and After Elimination of Outlying Interviewers

Description of Statistic	All Interviewers: Mean or Proportion	All Interviewers: p^*_{int}	After Elimination of Outlying Interviewers: Mean or Proportion	After Elimination of Outlying Interviewers: p^*_{int}
Two week doctor visits (from Person Section) None	.849 (1918)[a]	.0092*	.853 (1882)	.0033
Physician visits (from Supplements) None	.831 (1918)	.0081*	.842 (1798)	-.0054
Mean number of chronic conditions	.516 (1918)	.0097*	.518 (1800)	.0034
Health status Excellent	.418 (1918)	.0085*	.419 (1818)	.0001
Time since last dental visit 2wks. - 6months	.331 (1918)	.0018	.328 (1824)	-.0036

[a] number of interviews

There are at least three possibilities that explain this phenomenon: 1) On these measures most interviewer behavior will produce similar means for their respondent groups, but a small number of interviewers will depart from the survey mean greatly. Thus, we would expect a similar finding in replications of this design. 2) The one or two outliers are not expected in repetitions of the survey. They represent cases in the tails of the distribution for interviewer means and are unlikely to be found in another survey. Thus, a better estimate of the intraclass correlation due to interviewers is obtained by deleting the outlying cases. 3) The outliers are interviewers with low response rates (or very different response rates from most) and thus they are attributable to a confounding of nonresponse bias and response error.

We cannot test these various hypotheses without a replication of the survey, but we can note that. 1) the identities of the outlying interviewers vary over measures. That is, the same interviewers are not consistently outliers on all measures. 2) For that reason, the outliers are not uniformly those with higher or lower response rates. This variation over measures in the identity of outlying interviewers appears to dismiss effectively the hypothesis of nonresponse bias explaining the outliers. It also threatens the speculation that this pattern would not occur in replications of the survey, since the variability is not a function of only one or two interviewers.

6. Correlates 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 the interviewers as the unit of analysis. At that level of aggregation we can attempt to discover correlates of variability in interviewer means or in the deviation of individual interviewer means from the overall study value. We have hypothesized that this deviation is related to interviewer behavior and that we can measure this behavior through a monitoring process.

To examine the relationship between monitored behavior and variability, we chose to look at the five statistics in Table 3 with large values of p^*_{int}. For each of these variables we created scatterplots of the squared deviation of the individual interviewer's means from the study mean by two monitoring variables, proportion of time the question was read correctly and proportion of time the question was read well. Thirty of the thirty-three available interviewers were monitored over several occurrences of each question.

All ten scatter plots were similar in appearence. We expected an inverse relationship between the size of the individual interviewer's squared deviations and the proportion of "correct" behavior. After examining the plots we concluded that there was no apparent relationship.

Since monitored behavior did not prove to be a good predictor of interviewer variability, we considered other interviewer characteristics that are thought to measure performance. Response rate, size of workload, hours per interview and number of hours worked on the study were plotted against the interviewer's squared deviations on the five dependent variables previously described. Again we found no apparent relationship between any one of these variables and interviewer deviations. The so called "better" interviewers did not deviate any more or any less from the overall mean than did the other interviewers.

In one last attempt to explain the variability found among interviewer means we examined the mean value of interviewers' squared deviations for categories of several variables used to evaluate each interviewer's performance. These included such things as cooperation, efficiency, commitment to quality and standards question asking ability, speech and pace, and eliciting respondent participation. For each variable, interviewers were rated on a five point scale ranging from poor to excellent, by their supervisor. In all we examined 13 such variables, again hypothesizing that interviewers with poor performance ratings would have larger deviations. This was not the case, no pattern of interviewer rating and size of deviation emerged.

7. Summary

Our attempts to study the nature of variation in response errors across interviewers has not been completely successful. This lack of success, we think, forms a challenge to designers of interviewer training activities and reflects unusually low measured interviewer effects The monitoring procedures used were measures of the adherence of the interviewers to the behavior prescribed by the survey designers. These training procedures, in turn, were developed using beliefs about ways of reducing both interviewer variance and interviewer bias. The monitoring data produced generally high ratings of interviewers, but lower than those in our previous use of the procedures. That study contained generally higher values of ρ_{int}, but better interviewer ratings on the monitoring data. The variation in monitoring ratings of interviewers seemed, however, to have little relationship with the variation in interviewer means. Our attempts to examine this covariation is hampered by the generally low magnitudes of interviewer effects. Two possible interpretations could be made: 1) the variation in response errors across interviewers is not systematic (i.e , we would not obtain the same findings in repeated trials), 2) the monitoring data do not reflect the true sources of interviewer variation , there are other characteristics of interviewer behavior that produce different expected response deviations across respondents. We have no way of rigorously testing these alternative hypotheses, but the lack of relationship between the monitoring data and the response data for the second time raises suspicions that the measurements obtained by the monitors are not relevant to response deviations associated with interviewers. If this is true the monitoring procedures should be altered to measure adherence to other aspects of the training guidelines or the hiring and training procedures for interviewers need revision to address the causes of interviewer variation.

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.

Biemer, P., The Estimation of Non-Sampling Variance Components in Sample Surveys, Unpublished Ph.D. thesis, Texas A&M University, 1978.

Cannell, C. and Fowler, F., "A Study of Reporting of Visits to Doctors in the National Health Survey," Ann Arbor, Mich.: Survey Research Center, The University of Michigan, 1963.

Cannell, C. Lawson, S., and Hausser, D., A Technique for Evaluating Interviewer Performance, Ann Arbor, Mich.: Survey Research Center, The University of Michigan, 1977. Code No. 4022.

Feather, Joan, "A Study of Interviewer Variance," WHO International Collaborative Study of Medical Care Utilization, Saskatchewan Study Area Reports, Series II, Monograph No. 3, 1973.

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

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

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

Groves, R.M. and Magilavy, L.J., "Estimates of Interviewer Variance in Telephone Surveys," Proceedings of the American Statistical Association, Section on Survey Research Methods, 1980, 622-627.

Hansen Morris H. Hurwitz, William N. and Bershad, Max A., "Measurement Errors in Census and Surveys," Bulletin of the ISI, 38 Volume No. 2, 1961. 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, 1958, 635-655.

Kish, Leslie, "A Procedure for Objective Respondent Selection Within the Household," Journal of the American Statistical Association 44, 1949, 380-387.

Lansing, J.B., Withey, S.B. and Wolfe A.C. Working Papers on Survey Research in Poverty Areas, Survey Research Center, The University of Michigan, 1971, pp. 219-318.

Marquis, K.H. and Cannell C F., A Study of Interviewer-Respondent Interaction in the Urban Employment Survey, Survey Research Center The University of Michigan, 1969.

Mathiowetz, N.A. and Cannell, C.F., "Coding interviewer Behavior as a Method of Evaluating Performance," Proceedings of the American Statistical Association, Section on Survey Research Methods, 1980.

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

This project has been funded at least in part with Federal funds from the Department of Health and Human Services. The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services. The Survey Research Center staff having major responsibility for this research includes. Charles Cannell, Robert Groves, Lou J. Magilavy, Nancy Mathiowetz, and Peter Miller.