DOCUMENT RESUME

ED 271 437 SP 027 742

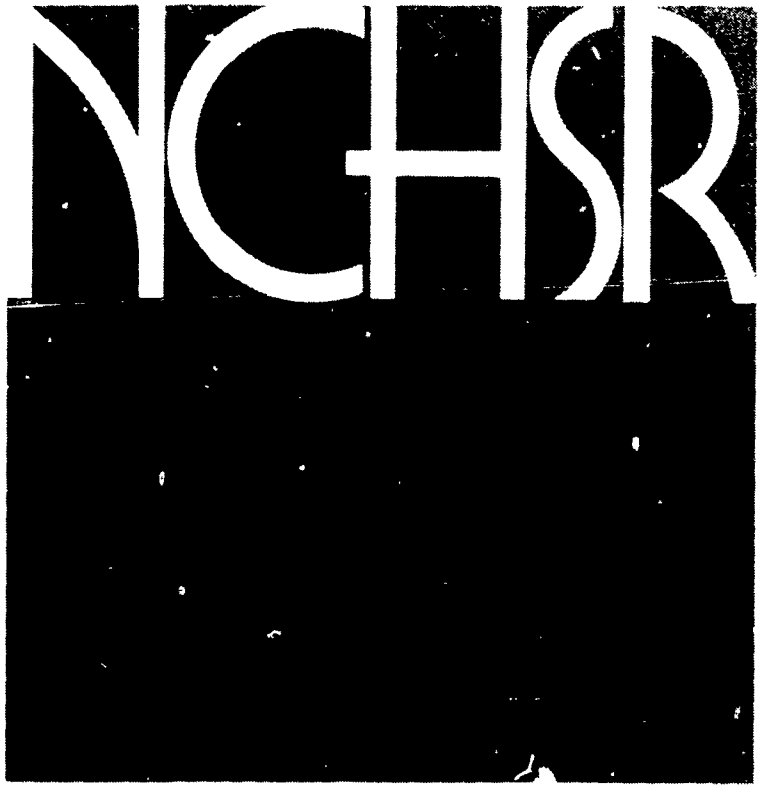
AUTHOR Fowler, Floyd J., Jr.; Mangione, Thomas W.
TITLE Reducing Interviewer Effects on Health Survey Data. Executive Summary.
INSTITUTION National Center for Health Services Research and Health Care Technology (DHHS/PHS), Rockville, MD.
SPONS AGENCY National Center for Health Services Research and Development (DHEW/PHS), Rockville, Md.
PUB DATE Feb 86
GRANT 3-R18-HS04189
NOTE 22p.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Data Collection; Health Services; Information Seeking; Inquiry; Interrater Reliability; *Interviews; *Job Training; Questioning Techniques; *Training Methods

ABSTRACT

This large-scale field experiment examined the potential of various training and supervision programs to affect the performance of health survey interviewers and the quality of data they collect. It was found that interviewers who received less than one day of basic training generally displayed inadequate interviewing skills. A program of tape recording as part of the supervision of household interviewers was associated with more precise and less biased data if interviewers were more than minimally trained. Training and supervision were found not to be compensatory but, rather, to interact so that if either was inadequate the data were adversely affected. The results also point to the value of designing questions to minimize the need for probing, a significant source of interviewer effects, and the value of procedures to communicate the importance of accuracy to respondents. Overall, attention to a variety of aspects of interviewer management--their training and supervision, the design of questions, the procedures they are to use, and the size of their assignments--are cost-effective ways to improve the quality of survey-based estimates. (Author)

# Reducing Interviewer Effects on Health Survey Data




NATIONAL CENTER FOR HEALTH SERVICES RESEARCH AND HEALTH CARE TECHNOLOGY ASSESSMENT





U S DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service

## Foreword

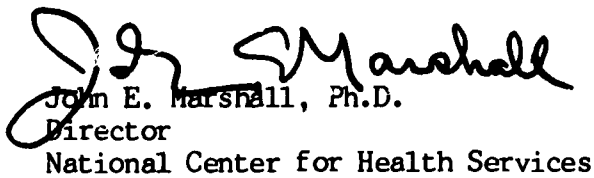
This report is a summary of research conducted by Floyd Fowler and Thomas Mangione on the costs and benefits of increased interviewer training and supervision. It fills a major void in current knowledge about the magnitude and nature of non-sampling error in health surveys. Their study is one of the first to provide more than anecdotal evidence about: (1) the optimum length of interviewer training, (2) the best method of supervision and (3) the implications of the size of interviewer assignments. In addition, it adds to previous knowledge about questionnaire design and wording.

The major value of this report lies in the set of practical recommendations that are made to improve the quality of health survey data. They include: (1) increase interviewer training beyond one day, (2) write questions to minimize the need for interviewer probing, (3) tape-record all interviews or a sample of interviews for supervisory review, and (4) reduce the size of interviewer assignments. All four strategies for reducing interviewer effects are quite cost-effective relative to increasing sample size, the most common approach to increasing the precision of survey estimates.

Since this study shows that the relationship between the length of training and data quality is not linear, the finding that one day of training is not adequate should not be interpreted to mean that "more is better." A complex interaction between length of training and mode of supervision was found that suggests that too much training may even be counter-productive without intensive supervision. In addition, it must be recognized that the optimum length of training is a very study-specific issue that will depend, to a large degree, on the complexity of the instrument.

An especially valuable contribution is made from the finding that taping interviews is a very cost-effective alternative to the usual method of direct supervisory observation. This finding should have a significant impact on the conduct of future health surveys. Finally, for the same level of precision, smaller interviewer assignments and hence larger staffs would be less costly than fewer interviewers taking larger assignments.

The results of this study should enable survey planners and researchers to make more informed decisions concerning the tradeoffs that affect data quality in health surveys. By presenting clear and easy-to-implement methods of reducing survey costs while increasing data quality, Fowler and Mangione have made an important methodological contribution to the health services research community.



John E. Marshall, Ph.D.
Director
National Center for Health Services Research
and Health Care Technology Assessment
February 1986

# EXECUTIVE SUMMARY

# REDUCING INTERVIEWER EFFECTS ON HEALTH SURVEY DATA

by

Floyd J. Fowler, Jr.
and
Thomas W. Mangione

Center for Survey Research
University of Massachusetts/Boston

Revised
February, 1986

Supported under Grant # 3-R18-HS04189
from the
National Center for Health Services Research

TABLE OF CONTENTS

ABSTRACT.....1

BACKGROUND.....2

OVERVIEW OF METHODS.....3

THE EFFECTS OF INTERVIEWER TRAINING AND SUPERVISION.....4

OTHER WAYS TO AFFECT DATA QUALITY.....8

Interviewer Procedures.....8

Question Design.....10

Size of Interviewing Staff.....10

CONCLUSION.....11

REFERENCES.....14

APPENDIX.....15

ABSTRACT

This large-scale field experiment examined the potential of various training and supervision programs to affect the performance of health survey interviewers and the quality of data they collect. It was found that interviewers who received less than one day of basic training generally displayed inadequate interviewing skills. A program of tape recording as part of supervision of household interviewers was associated with more precise and less biased data if interviewers were more than minimally trained. Training and supervision were found not to be compensatory but, rather, to interact so that if either was inadequate the data were adversely affected. The results also point to the value of designing questions to minimize the need for probing, a significant source of interviewer effects, and to the value of procedures to communicate the importance of accuracy to respondents. Overall, attention to a variety of aspects of interviewer management - their training and supervision, the design of questions, the procedures they are to use, and the size of their assignments -- are cost-effective ways to improve the quality of survey-based estimates.

## BACKGROUND

Many areas of social policy rely heavily on data developed by survey research techniques for planning and evaluation purposes. Obviously, the quality of estimates derived from surveys is of considerable importance in the effective pursuit of social policy goals.

Methodologists often cite three sources of error, or reasons why figures derived from a sample survey do not accurately describe the true values for the populations from which they were drawn:

A. Sampling error, the normal chance variability that occurs because a sample may differ within a calculable range from the population from which it was drawn.

B. Nonresponse error, error resulting from the fact that data are not collected from every population member chosen to be in a sample.

C. Measurement or response error, error stemming from the fact that answers to questions do not perfectly measure what the researcher was trying to measure. Factors that affect response error include problems with the questions and the way they are designed, limitations on a respondent's ability and willingness to answer questions accurately, and problems with the way that an interviewer handles the question and answer process.

Current thinking about the design and execution of surveys emphasizes "total survey design". By this methodologists mean that researchers should take into account all potential sources of error when designing a survey or evaluating survey data. Although such a view may seem only reasonable, for practical reasons it has not been common practice in the past.

Sampling error has long been a concern of sampling statisticians. The limits on the confidence one can have in estimates from a random sample of a particular size and design can be readily calculated. Most reports of survey estimates include some acknowledgement of the role of sampling error in the precision of the figures and usually attach some numerical estimates of sampling error.

The significance of nonresponse for survey estimates is also commonly acknowledged. Although the effect of nonresponse on survey estimates seldom can be calculated very well, most researchers attempt to achieve a respectable response rate and they commonly report the rate of response.

In contrast, other aspects of the data collection that directly affect the quality of the measurement process are frequently overlooked entirely both in the design of surveys and in the reporting of the data collection. For example, the amount and kind of effort that went into the development of the questions, and evidence for the validity of the answers, are only rarely reported. Most germane to the current topic, however, although their importance is well documented (e.g. Hyman, 1954), consideration of the interviewers and the quality of interviewing in a survey is typically almost totally ignored.

Interviewers have three different roles to play in a sample survey. First, they are the ones who implement the sample design. How they do that affects response rates and costs.

The second role of the interviewer is to train and motivate respondents to do their part in the interview.

The third job of the interviewer is to handle his or her side of the question and answer process. Specifically, the interviewer asks the questions, clarifies questions as needed, stimulates the respondent and directs his or her effort in the event that an initial answer does not meet question objectives, and records the answers.

We know from previous research that interviewers can influence the quality of estimates in two different ways. First, if interviewers are not consistently standardized, survey-based estimates are less precise than they otherwise would be for a sample of a given size. In essence, lack of standardization increases the amount of random error around the survey estimates and decreases the extent to which true differences among respondents are detectable in their answers.

Interviewers can also systematically bias data or make them less valid. Cannell (1977a and 1977b) has demonstrated that the pace at which an interview is conducted, the kind of respondent behaviors that interviewers reinforce during the interview, and the goals that interviewers communicate to respondents can all be related to the accuracy with which respondents report.

Although methodological studies have left little doubt that interviewers have a role to play in the quality of survey estimates, in a typical survey, the effects of the interviewers cannot be dissociated from other sources of error. Except for the response rate, for which interviewers have some responsibility, the effects of poor interviewing typically are not observable in data. Moreover, researchers have not had good information about the costs and potential benefits of various strategies to improve the quality interviewing. This lack of information made it difficult for researchers realistically to consider trade-offs between the quality of interviewing and such design decisions such as the size of the sample or the response rate.

The research reported here was an explicit attempt to provide the information that researchers need about the role of interviewers in affecting survey data and the benefits of various options available to researchers to improve the quality of survey data.

## OVERVIEW OF METHODS

This study was designed to assess the value of various realistic options for training and supervising interviewers for improving the quality of survey data. The study also was designed to provide data about the properties of questions that make them susceptible to interviewer effects.

Four training programs were examined. The shortest program lasted only about half a day. The longest training program tested was ten days long, which is considerably more intensive and extensive than any training program routinely used by survey research organizations. The other two programs,

lasting two and five days respectively, are typical in length to training programs commonly in use.

With respect to supervision, three levels were tested. The minimal program provided interviewers with only feedback on costs and response rates. Level II added review of a sample of completed interview schedules; Level III involved tape recording all interviews and providing feedback on interviewing techniques, as well as costs, response rates, and the quality of completed interviews.

In all, 57 newly hired interviewers were randomly assigned in a balanced design; first to one of the four training programs, then to one of the three supervision programs. They carried out a special purpose health interview; on average they each took 26 interviews.

There were four features of the experiment that we considered essential to its value and success. First, this large-scale project was designed to have enough power to detect real effects and reach defensible conclusions. Second, in order to permit study of the effects of interviewers on data and evaluate the quality of their work, it was necessary to dissociate interviewers from idiosyncrasies of their sample. To do this, each interviewer's sample of respondents was representative of the sample population as a whole. Third, it was important to generalize about question form and content and how they relate to interviewer effects. Therefore, the health survey questionnaire was carefully constructed to include an array of common questions, as well as a sample of various types of questions.

Fourth, we wanted to know not only whether or not the experimental training and supervisory programs affected data, but also to understand the whys, and to gain a more general understanding of the role of the interviewer in the data collection process. In order to do this, two special additional data collection activities were built into the project: 1) all respondents who were in the survey were also reinterviewed to gain information about the respondent's reaction to the survey process and to the interviewer; 2) information was collected directly from interviewers after they were finished with their assignments about their perceptions of the job and of the interview process.

These data, in combination with information derived from coding the tape recorded interviews taken by Supervision Level III interviewers, provided a unique opportunity to study what interviewers actually do and how they affect data.

## THE EFFECTS OF TRAINING AND SUPERVISION

The central hypothesis tested in this study was that more extensive programs of training and closer supervision of interviewers would improve the quality of data that they collected. To examine this hypothesis, two measures of data quality were created.

First, one goal of good survey interviewing is that interviewers be standardized; that is, that they do not affect the answers that they obtain. The effect of lack of standardization is a reduction in the precision of survey estimates. A measure of the extent to which answers can be predicted by knowing the interviewer, and hence one can infer answers were affected by

the interviewer, is *rho*, the intraclass correlation, as proposed by Kish (1962).

The validity of data, in the absence of a credible criterion, is more difficult to evaluate. However, there are some questions for which we are able to make a good guess about the direction in which reported data are most likely to differ from the true answers. One such kind of question involves reporting the number of events occurring over a period of time. For many such questions, underreporting has been documented. Another class of questions, those that may have socially desirable answers, have been shown to be prone to error; overall people tend to err in the direction of making their answers more attractive or socially acceptable than the true value. Based on these premises regarding patterns of error or bias, survey questions in the study were selected for which a direction of "better", or less biased data could be specified. Because each interviewer's sample of respondents was a random subsample of the total sample, any differences between the average answers given by an interviewer's respondents and those given by other interviewers' respondents could be attributed to the effects of the interviewer rather than to sample differences.

Tables 1 and 2 present the results of analyses of variance looking at the relationship between the training and supervision programs to which interviewers were assigned and the values of these two measures of the quality of data they collected. A number of important observations arise from these two tables.

1) Training and supervision do matter. For both measures of quality, the effect of the combination of training and supervision received was statistically significant.

2) Tape recording interviewers in order to provide direct supervision of the way they handle the question-and-answer process in the interview improves the precision of survey estimates and, if interviewers have had more than minimally adequate training, also probably improves the validity of the data they collect.

3) Those who received the most training and were tape recorded were, as a group, the best interviewers overall. Their data were significantly less biased than the other groups, and their level of standardization was equal to, or better than, any other group.

4) There is a complex interaction between training and supervision. Although we hypothesized that they might be complementary, with, for example, more training compensating for minimal supervision, or vice versa, our findings are quite different. Instead we found that if either was inadequate, the quality of the data was diminished. Specifically, well trained but poorly supervised interviewers were among the worst performers on both quality dimensions, while intensive supervision of the least trained interviewers did improve their degree of standardization but also produced data that were more biased.

The study was designed not only to find out whether training and supervision affected data, with an eye to setting some minimal standards in those areas, but also to understand better the ways in which training and

TABLE 1

AVERAGE RHO (X1000) BY LEVEL OF TRAINING AND SUPERVISION*

| Length of Training Program | Supervision Level I | Supervision Level II | Supervision Level III | Average |
|---|---|---|---|---|
| < 1 day | 14 | 10 | 8 | 11 |
| 2 days | 12 | 6 | 11 | 9 |
| 5 days | 9 | 9 | 8 | 9 |
| 10 days | 15 | 20 | 7 | 12 |
| Average | 12 | 10 | 8 | 10 |

*******

| Analysis of Variance | df | Sum of Squares | F | P |
|---|---|---|---|---|
| Training | 3 | 22.19 | 1.94 | .11 |
| Supervision | 2 | 14.60 | 1.92 | .15 |
| Interaction | 6 | 47.67 | 2.09 | .05 |

| | |
|---|---|
| Mean Square Model | 7.68 |
| df | 11 |
| Mean Square Error | 3.81 |
| df | 648 |
| F | 2.02 |
| P | <.05 |

Contrasts**

Supervision Level III (taped) vs. I & II (not taped) t = 1.6

*Analysis includes only items for which interviewers affected the answer ($p<.10$ by F test). Rho was transformed to Log $\frac{\text{Rho}}{1\text{-Rho}}$ prior to analysis to more closely meet assumptions of normality.

**Probabilties indeterminate because multiple contrasts were run, but t values meet or exceed usual values for 1-tailed test of significance ($P<.05$)

TABLE 2

AVERAGE STANDARD SCORE (X100)* ON QUESTIONS
JUDGED MOST LIKELY SUBJECT TO SYSTEMATIC BIAS
BY LEVEL OF TRAINING AND SUPERVISION

| Length of Training Program | Supervision Level | | | |
|---|---|---|---|---|
| | Level I | Level II | Level III | Average |
| < 1 day | 18 | 11 | -26 | 1 |
| 2 days | 7 | -5 | 43 | 15 |
| 5 days | -8 | -5 | 7 | -2 |
| 10 days | 0 | 21 | 52 | 27 |
| Average | 4 | 5 | 20 | 10 |

*******

| Analysis of Variance | df | Sum of Squares | F | P |
|---|---|---|---|---|
| Training | 3 | .49 | 2.44 | .06 |
| Supervision | 2 | .18 | 1.36 | .26 |
| Interaction | 6 | 1.06 | 2.64 | .01 |

| | |
|---|---|
| Mean Square Model | .157 |
| df | 11 |
| Mean Square Error | .067 |
| df | 3623 |
| F | 2.35 |
| P | <.01 |

Contrasts**

| | |
|---|---|
| Taped vs. not taped (excluding 1-day training) | t = 1.77 |
| 10-day training and taped vs. rest | t = 1.92 |

*A positive score is a score judged to be less biased.

**Estimates were adjusted for the fact of multiple measures per interviewer. Probabilities indeterminate, because multiple contrasts were run, but t values exceed usual values required for 1-tailed test of significance (P<.05).

supervision matter for interviewers, and the ways in which interviewer behavior affects data.

One analysis related interviewer training to the specific quality of their interviewing skills. In particular, we looked at the quality of asking questions as worded, the quality of probing, and recording answers appropriately, and maintaining a neutral interpersonal relationship. For the most part, our conclusions were based on coding of tape recorded interviews.

When we looked at these basic interviewing skills in relationship to training, it was quite evident that giving interviewers' less than one day of training resulted in inadequate interviewing skills compared to the other interviewers (Table 3). These are highly significant effects. Beyond that, however, the increments in basic interviewing skills demonstrated by interviewers with increasingly long training programs tended not to be significant, with the exception of probing skills.

One might have thought that an effect of intensive supervision of interviewing behaviors would be to improve interviewer skills over time. This proved not to be the case. In fact, the basic interviewing skills of intensively supervised interviewers did not improve in the second half of their work compared to their first half. However, we did find evidence that skills of interviewers who were not tape recorded deteriorated in the second half of their assignments.

Thus, these data provide a basis for understanding the results of Tables 1 and 2. Interviewers need at least a couple of days of training to gain skills that produce standardized data collection. Intensive supervision through review and feedback from taped interviews maintains these skills that produce standardized data. Interviewers who were not intensively supervised showed deterioration in skills, particularly among those who acquired higher skill levels to begin with.

However, intense supervision did not create skills that were not acquired from training. Intense supervision of poorly trained interviewers produced nervous interviewers who gathered biased data. It is most important to note, however, that those who received the most training and were intensively supervised gathered the best data on both dimensions of quality.

## OTHER WAYS TO AFFECT DATA QUALITY

Although the study was organized to look at the way training and supervision affect interviewer behavior and data quality, there are at least three other aspects of interviewer management that we came to appreciate as being important in reducing the contribution that interviewers make to data error. First, procedures that interviewers are asked to use in carrying out the interview affect quality. Second, the way questions are designed affects the likelihood that interviewers will affect the answers. Third, the size of the interviewers' assignments affects the potential impact of interviewers on the quality of data.

### Interviewer Procedures

In the field of survey research, there is a commonly accepted set of guidelines for interviewers regarding asking questions as worded, probing in a

TABLE 3

SELECTED MEASURES OF INTERVIEWER BEHAVIOR
FROM CODING TAPED INTERVIEWS BY TRAINING PROGRAM
(SUPERVISION LEVEL III ONLY)

| Interviewer Behaviors from Tape Coding | Length of Training Program | | | | P |
|---|---|---|---|---|---|
| | < 1 day | 2 days | 5 days | 10 days | |
| Average No. Questions Read incorrectly/ Interview | 21 | 7 | 14 | 6 | <.01 |
| Average No. of Directive Probes/Interview | 8 | 5 | 5 | 3 | <.01 |
| Average No. of Times Failed to Probe Inadequate Answers/Interview | 8 | 6 | 5 | 5 | <.01 |
| Average No. of Inaccurate Recording of Closed Question Answers/Interview | 1 | 1 | 1 | * | .05 |
| Average No. of Inaccurate Recording of Open Question Answers/ Interview | 4 | 2 | 2 | 2 | <.01 |
| Average No. of Instances of Inappropriate Feedback/Interview | 2 | * | * | * | <.01 |
| Percentage of Interviews Rated Excellent or Satisfactory | | | | | |
| Reading Questions as Worded | 30 | 85 | 72 | 84 | <.01 |
| Probing Closed Questions | 48 | 67 | 72 | 80 | <.01 |
| Probing Open Questions | 16 | 44 | 52 | 69 | <.01 |
| Recording Answers to Closed Questions | 88 | 88 | 89 | 93 | .74 |
| Recording Answers to Open Questions | 55 | 80 | 67 | 83 | <.01 |
| Non-biasing Interpersonal Behavior | 66 | 95 | 85 | 90 | <.01 |

* Less than 0.5 times per interview.

nonbiasing fashion, not influencing the answers through the recording process, and being interpersonally neutral. Our analyses support the notion that good question reading and non-directive probing are basic skills for reducing interviewer effects.

With respect to bias, if we were to advocate one characteristic of an interviewer for obtaining good data, it would be that he or she conveys to the respondent that the accuracy of the data is important. Interviewers whom respondents rated as being most concerned about accuracy produced significantly less biased data. Specific ways that interviewers can communicate the importance of accuracy include the pace at which the interview proceeds and being attentive to the interviewer's own role, by asking questions exactly as worded and probing carefully.

In addition, there were some indications in our data that being friendly and relating to respondent needs may play some role in the accuracy of data collected. There was a correlation between respondent-rated friendliness and our measure of bias among the tape recorded interviewers ($r = .31$).

In terms of how to achieve these goals, in addition to simply telling interviewers what to do, our findings tended to reinforce the salience and relevance of the techniques Cannell (1977b) tested - giving interviewers specific instructions in how to train respondents and reinforce accurate reporting as a goal.

Question Design

As others have found, it was apparent that interviewers had more trouble with some questions than others. Our results were consistent with results reported by Groves and Kahn (1979) showing a fourth to a third of survey items were subject to significant interviewer effects.

The specific questions that were most subject to interviewer effects were the ones that interviewers had to probe frequently. When interviewers are required to probe in order to obtain an adequate answer, it produces opportunities for them to be inconsistent, to use directive probes, or to fail to probe inadequate answers. This finding is of considerable practical significance because better and more systematic pretesting of survey questions can identify problem questions. By rewriting these items we can probably reduce the need for probing and thereby increase the precision of estimates.

The most pervasive hypothesis in the literature about item types is probably that sensitive questions may be most subject to interviewer effects. We found just the opposite, with respect to precision of estimates (not bias itself) which intriguingly is similar to findings reported by Bradburn and Sudman (1979). We found that interviewers were more consistent in the way they handled sensitive questions than they were for the average question. This led to lower interviewer effects for sensitive questions.

Size Of Interviewing Staff

Another dimension of a study design to which a researcher could attend in order to reduce error is the number of interviews taken by an interviewer. The total effect of interviewers on the precision of estimates in a study is a product of the intraclass correlation for interviewers and the average number

of interviews taken per interviewer. For a given level of intraclass correlation, smaller numbers of interviews per interviewer (and hence a larger interviewer staff), will produce more precise data.

In addition, we have suggestive evidence that the quality of data deteriorates as more interviews are taken by an interviewer. This is similar to results reported by Cannell (1977a).

Thus, from the point of view of reliability, as reflected in intraclass correlations and from the point of view of validity, it seems as if using more interviewers, thus having each interviewer take fewer interviews on average, is a constructive way to improve the quality of survey data. This recommendation, of course, assumes a similar level of supervision and training.

## CONCLUSION

Good methodology has been defined as designing a study to get the most precision or accuracy for a dollar. To do this, one should attend to the various features of any data collection enterprise that affect the data that result.

For too long, only the calculation of sampling errors and response rates has passed for methodological rigor. Features of survey design that may be equally or more important in the overall quality of survey estimates, notably survey question design and interviewer performance, often are ignored.

Of course, it is easy to understand why these considerations can be ignored; the effects of poor question design and of poor interviewing are not immediately apparent when data are analyzed. It takes special calculations to find out how well the measurement process has actually been carried out. Yet, the fact that error is not obvious does not mean the error is not there.

There are a good number of questions (probably more than half in most health surveys) that are relatively unaffected by interviewers. However, about a fourth to a third of the questions in representative health surveys are significantly influenced by the quality of interviewing. These are not obscure or unimportant questions. Some commonly used questions which were subject to significant interviewer effects include:

| | *Rho* |
|---|---|
| Are you limited in any way because of a disability or health condition? | .014 |
| In the past 12 months, did you have hemorrhoids or piles? | .017 |
| In the past 12 months, did you have deafness in one or both ears? | .020 |
| How many days in the last month would you say you had (USUAL NUMBER) of drinks? | .034 |
| How long ago was the last time you were actually seen by a doctor about your health - within the last month, 1 to 6 months ago, 6 months to a year ago, or more than a year ago? | .037 |

When this project began, we knew that interviewers had a largely unappreciated role in the creation of survey error. Moreover, we knew that there were very few practical guidelines for researchers and for those who would purchase research for how to minimize the effects of interviewers on survey data. In the preceding pages, we have outlined five very practical and useful ways, in addition to increasing sample size and minimizing nonresponse, by which researchers can improve their estimates by improving the way that interviewers do their jobs. Although the value of stressing accuracy to respondents emerges from our data, the techniques for doing this were developed in Cannell's work (e.g. 1977b). However, this project clearly supports the value of four additional strategies aimed at improving survey estimates by reducing interviewer effects on data.

Table 4 summarizes these four strategies, along with the most common approach to increasing the precision of survey estimates, increasing the size of the samples.

1. Give more than minimal training in basic interviewing skills.

2. Make tape recording, review, and feedback a standard part of supervision.

3. Try to design questions to make asking them easy and reduce the need for probing, and pretest them thoroughly to make sure the attempt is successful.

4. Reduce the number of interviews taken per interviewer.

Of course, the cost effectiveness of the steps outlined in Table 4 and how much estimates will be improved will vary from setting to setting and estimate to estimate. However, for that important subset of items that interviewers significantly influence, steps to reduce interviewer effects are quite cost effective ways to improve the quality of estimates. We believe these data provide practical guidelines for researchers for improving their survey data by attending to the quality of interviewing. Standards for the way interviewers are managed have been too long absent, despite a history of research showing that interviewers matter. We are hopeful that a concrete effect of this project will be to help bring attention to the interviewer as part of the total design of surveys to the status it deserves.

TABLE 4
FIVE WAYS TO DECREASE STANDARD
ERRORS OF ESTIMATES

| | Approach | Likely Cost | Effect on Standard Errors |
|---|---|---|---|
| Sample Size[1] | Increase effective sample size by about 20% | About a pro rata increase in data collection and data reduction costs | Decrease by 10% |
| Interviewer Training[2] | If interviewers receive less than 1 day of basic training, increase by a day or two | Equivalent to about 12 hours of interviewer wages per extra training day per interviewer | Decrease by 10% for the 1/3 of survey items which are most affected by interviewers |
| Tape Supervision[3] | Tape all or a sample of interviews, review one a week per interviewer provide feedback | About 2 hours/ interviewer per week | Decrease by more than 10% for 1/3 of items most affected by interviewers |
| Question Design[4] | Rewrite questions to reduce need for probing and make administration and reading of questions easier | About twice the length of the interview to tabulate interviewer behavior from taped pretest interviews plus time to rewrite questions | Efficacy not yet demonstrated but data suggest noticeable gains likely |
| Number of Interviews Per Interviewer[4] | Reduce assignment size by 20% by using 20% more interviewers | Difficult to estimate but certainly less than changing sample size to produce same effect | Decrease by 10% for 1/3 of items most affected by interviewers |

NOTES:

1. If complex design rather than simple random sample, may entail more than 20 percent more interviews.
2. Clearly produces direct effects on standard errors only beyond minimal training. However, even more training may also pay off in decreasing bias in data.
3. Probably even greater benefits over time as interviewers deteriorate without taping and feedback. Also, significantly reduces bias for adequately trained interviewers.
4. Probably also reduces bias through reduced burn out.

REFERENCES

Bradburn, N. A., Sudman, S. and Associates (1979) Improving Interview Methods and Questionnaire Design, San Francisco: Jossey-Bass.

Cannell, C. F., Marquis, K. H. & Laurent, A. (1977a) "A summary of studies." Vital and Health Statistics, Series, 2, 69. Washington. DC: U.S. Government Printing Office.

Cannell, C. F., Oksenberg, L., & Converse, J. M. (1977b). Experiments in Interviewing Techniques: Field Experiments in Health Reporting: 1971-1977. Hyattsville, MD: NCHSR.

Groves, R. M., & Kahn, R. L. (1979) Surveys by Telephone: A National Comparison with Personal Interviews. New York: Academic Press.

Groves, R. M., & Magilavy, L. J. (1980) Estimates of interviewer variance in telephone surveys. Proceedings of the American Statistical Association, Survey Research Methods Section, 622-627.

Hyman, H. H., Cobb, J., Feldman J., & Stember, C. (1954) Interviewing in Social Research. Chicago: University of Chicago Press.

Kish, L. (1962) Studies of interviewer variance for attitudinal variables. Journal of the American Statistical Association 44, 380-387.

## APPENDIX

### Calculation of *Rho* (Interviewer Effects)

One common barrier to attending to interviewer effects is the lack of a measure. When interviewers are assigned to samples purely on the basis of proximity (or some similar nonrandom criterion) *rho* cannot be meaningfully calculated. However, if interviewers have assignments that are random subsets of the whole sample, then interviewer effects can be calculated for each item in a survey. Also, for assignments that are roughly random (e.g. in centralized telephone facility studies), these calculations provide imperfect but useful estimates of interviewer effects.

A second barrier to these calculations is the availability of appropriate software. There are specialized programs to calculate interviewer effects, usually variations on programs designed to calculate sample design effects, but they are not available in many computer facilities. However, a standard analysis of variance program for Generalized Linear Models can be used to accomplish the same thing with reasonable precision.

For each interval or ratio scale variable run an analysis of variance with interviewer as the random effects variable. For nominal scale data, translate answers to dummy (i.e. 1/0) variables and use these as the dependent variables in the ANOVA. From the analyses of variance statistics note the Model Mean Square and the Error Mean Square terms and the average number of interviews per interviewer.

$$Rho = \frac{\dfrac{\text{Model MS - Error MS}}{n}}{\text{Error MS} + \dfrac{\text{Model MS - Error MS}}{n}}$$

### Calculation of Design Effects on Size of the Standard Error

$$\text{Design Effect} = \sqrt{1 + Rho\ (n - 1)}$$

n = average number of interviews per interviewer

If the number of interviews is reasonably similar, a simple average may be used. If they range widely, a more complex calculation is needed (see Groves and Magilavy, 1980).

Table A1 below provides sample calculations of design effects (defts) for various values of *rho* and interviewer assignment sizes.

TABLE A1

MULTIPLIERS OF ESTIMATES OF STANDARD ERRORS OF MEANS DUE TO INTERVIEWER EFFECTS* FOR SELECTED VALUES OF RHO AND AVERAGE INTERVIEWER ASSIGNMENTS

| Average Interviewer Assignment Size | Intraclass Correlation (Rho) | | | | |
|---|---|---|---|---|---|
| | .005 | .01 | .015 | .02 | .03 |
| 11 | 1.002 | 1.05 | 1.07 | 1.10 | 1.14 |
| 21 | 1.05 | 1.10 | 1.14 | 1.18 | 1.26 |
| 31 | 1.07 | 1.14 | 1.20 | 1.26 | 1.38 |
| 51 | 1.12 | 1.22 | 1.32 | 1.41 | 1.58 |
| 81 | 1.18 | 1.34 | 1.48 | 1.61 | 1.84 |
| 101 | 1.22 | 1.41 | 1.58 | 1.73 | 2.00 |

*Estimates of standard errors calculated from the sample size and design should be inflated by the multiplier in the table to take into account the effect of interviewers.

50272-101

| REPORT DOCUMENTATION PAGE | 1. REPORT NO. NCHSR 86-8 | 2. | 3. Recipient's Accession No. PB86-145745/AS |
|---|---|---|---|

**4. Title and Subtitle**
THE VALUE OF INTERVIEWER TRAINING AND SUPERVISION
Executive Summary

**5. Report Date** June 1985

**6.**

**7. Author(s)**
Floyd J. Fowler, Jr. and Thomas W. Mangione

**8. Performing Organization Rept. No.**

**9. Performing Organization Name and Address**
Center for Survey Research
University of Massachusetts/Boston
Boston, Massachusetts 02125

**10. Project/Task/Work Unit No.**

**11. Contract(C) or Grant(G) No.**
(C)
(G) HS 04189

**12. Sponsoring Organization Name and Address**
DHHS, PHS, OASH, National Center for Health Services Research and Health Care Technology Assessment (NCHSR)
Publications and Information Branch, 1-46 Park Building
Rockville, MD 20857 Tel.: 301/443-4100

**13. Type of Report & Period Covered**
1/1/81 - 3/31/85
Executive Summary

**14.**

**15. Supplementary Notes**

**16. Abstract (Limit: 200 words)**

This large-scale controlled experiment evaluated the extent to which interviewer training and supervision improves the quality of the data collected in health surveys. In addition, this study examined the kinds of interviewer behaviors that produce interviewer effects. Analyses have provided extensive knowledge about the role of the interviewer, in the context of total survey design, in producing error. Study findings suggest several strategies for reducing survey costs while improving data quality: (1) train interviewers for more than one day, (2) write questions so as to minimize the need for interviewers to probe, (3) tape-record all interviews or a sample of interviews for supervisory review, and (4) reduce the size of interviewer assignments. The results should enable survey planners and researchers to make more informed decisions concerning the tradeoffs that affect data quality in health surveys.

**17. Document Analysis a. Descriptors**

NCHSR publication of research findings does not necessarily represent approval or official endorsement by the National Center for Health Services Research and Health Care Technology Assessment or the U.S. Department of Health and Human Services.

Stephen E. Marcus, NCHSR Project Officer, 443-5780

**b. Identifiers/Open-Ended Terms**

Health services research, survey methods, interviewer training, data quality.

**c. COSATI Field/Group**

| 18. Availability Statement: | 19. Security Class (This Report) | 21. No. of Pages |
|---|---|---|
| Releasable to the public. Available from National Technical Information Service, Springfield, VA 22161 Tel.: 703/487-4650 | Unclassified | 17 |
| | 20. Security Class (This Page) Unclassified | 22. Price |

(See ANSI-Z39.18) See Instructions on Reverse OPTIONAL FORM 272 (4-77) (Formerly NTIS-35) Department of Commerce