

Barry L. Ford and Kathy Burstrom Hall, U.S. Department of Agriculture

ABSTRACT

This study was an exploration of the effect of four methods of data collection on agricultural estimates in seven states. These methods were: 1) data collection by using only personal interviews, 2) data collection by using telephone interviews and following up with personal interviews, 3) data collection by using mail questionnaires, following up with telephone interviews for mail inaccessibles, and then following up with personal interviews for telephone inaccessibles, and 4) data collection by using mail questionnaires and following up with personal delivery of questionnaires by enumerators to mail inaccessibles who were then asked to complete the questionnaires without the enumerators' help. The effects of the methods caused significant differences in the estimates of survey variables and nonresponse rates, and they exhibited significant interactions with the seven states.

INTRODUCTION

The U.S. Department of Agriculture currently uses three interviewing techniques -- mail, telephone, and personal interviews -- to collect data for surveys. The agency has always assumed that the three techniques have no impact on estimates but has never rigorously tested this assumption. To test the validity of the assumption, data on five variables in seven states were analyzed.

This study was general and exploratory in nature. Thus, the design of the project used a sample size which would only detect 10 percent differences in estimates. If gross differences were detected, then future research projects would be developed of a more specific and sensitive nature. The data for this study were divided into replicates in order to simplify the analysis. Details of the analytical methods are in the appendix.

The list sample in each of the seven states was systematically divided into four treatment groups: 1) a group receiving the current operational procedure of a sequence of mail, telephone, personal interviews (called the "operational" group in this report); 2) a group receiving a mail-personal delivery sequence; 3) a group receiving a telephone-personal interview sequence; and 4) a group receiving only personal interviews. Strictly speaking, the analysis of this report is not a comparison of interviewing techniques but a comparison of four data collection strategies -- the operational strategy and three strategies which are dominated by a specific interviewing technique. Any of these strategies can be applied to an entire sample. For example, although the third group is dominated by telephone interviews, a personal interview follow-up is necessary for that part of the group which can not be contacted by telephone. Thus, any of the four strategies is a procedure that could possibly be implemented into the operational program.

Data for the operational group were gathered by mailing a questionnaire to each farm operator who was sampled. Those operators who failed to respond within a few days were telephoned.

Whenever contact could not be made using the telephone, the questionnaires were sent to field enumerators who attempted to complete the reports by using any possible type of interview. Usually this meant a personal interview although sometimes operators were again telephoned.

The mail-personal delivery group employed a procedure that was the most different from the operational procedure. Farm operators received a questionnaire in the mail. If they failed to return it within the specified period of time, a field enumerator took them a copy of the questionnaire and asked the operator to fill it out without any assistance from the enumerator. The purpose of this procedure was to obtain data for the situation in which farmers read and interpret questionnaires without enumerator assistance as is the case in mail surveys. Although ESCS could implement the mail-personal delivery strategy as an operational procedure, that possibility was not the reason for making it one of the test procedures. The reason was to provide a control method which minimizes the effect of interviewers.

Data for most of the operators in the telephone-personal interview group were gathered by means of telephone interviews with calling conducted from the state office headquarters. Questionnaires for operators that were not contacted after repeated attempts or for which no telephone numbers were found were sent to field enumerators who tried to contact the operators in person.

Historically, the personal interview method has been a principal method of data collection for the U.S. Department of Agriculture, and currently it is the method used for most of its area frame surveys. During this study operators in the personal interview group were interviewed without any mailing or announcement of the survey immediately prior to the interview. The only use of the telephone was to set up appointments.

The survey used in this study is a hog survey run during each quarter of the year. A sample of farm operators, however, is used for an entire year. In all seven states except Wisconsin, the operators contacted during this project had also been contacted during the three previous quarters in order to answer survey questionnaires. These operators were familiar with the operational method. Therefore, they expected a telephone call or a personal interview when they did not respond to the mail questionnaires. Wisconsin used a new sample so operators were contacted for the first time.

This situation explains some of the problems associated with data collection during the project. The mail-personal delivery method proved to be the least popular with farmers. Several respondents preferred to be interviewed. Many asked the enumerators delivering the questionnaires why they had not been telephoned or were not being interviewed as in previous quarters. About one hundred operators simply refused to fill out the form themselves but would supply the information if they were interviewed. In these instances, enumerators conducted an interview rather than obtain no information at all. (This situation occurred most commonly when the farmers were

busy in the barn or the field when the enumerator arrived.) Although these problems were due primarily to a change in prior procedures, Wisconsin also reported similar problems which indicated that the personal delivery technique itself was unpopular. Thirteen of the operators contacted in the mail-personal delivery could not read. Despite some problems the mail-personal delivery method does present a contrast to the other methods in the percentages obtained by different types of interview. Table 1 gives the percentages of each method by type of interview.

TABLE 1
Type of Interview by Data Collection Methods
All Seven States

| Type of Interview Method | Completed by Mail | Completed by Telephone | Completed by an Enumerator Contact (Personal Interview or Personal Delivery) |
|------------------------------|-------------------|------------------------|--|
| Operational | % 25.3 | % 60.3 | % 14.4 |
| Mail-Personal Delivery | 34.3 | 1.6 | 64.1 |
| Telephone-Personal Interview | 0.0 | 90.2 | 9.8 |
| Personal Interview | 0.1 | 2.7 | 97.2 |

Nonresponse is a term used to combine two different situations. In one case, the farm operator refuses to supply the information requested and is classified a "refusal". In the other, the interviewer can not locate anyone who can supply the requested information, and the sample unit is classified as "inaccessible". Table 2 shows the response rates for each method. The results indicate the refusal rate for the personal interview method is much lower than for any of the other data collection methods. The impact of this result is discussed in the next section.

TABLE 2
Response Rates by Data Collection Methods
All Seven States

| Response Method | Completed Reports | Refusals | Inaccessibles |
|------------------------------|-------------------|-----------|---------------|
| Operational | % 87.2 | % 10.0 | % 2.8 |
| Mail-Personal Delivery | 86.5 | 10.6 | 2.9 |
| Telephone-Personal Interview | 88.2 | 10.0 | 1.8 |
| Personal Interview | 91.1 | 6.2 | 2.7 |

ANALYSIS

Mean values for four hog variables as well as the nonresponse rates for each of the four data collection methods are in Table 3. The columns labelled "All Data" represent the mean values of all respondents. The columns labelled "Positive Data" represent the mean values of those respondents who reported one or more hogs. The justification for examining positive reports

and omitting zero reports is that a response of "I don't have any hogs" provides little opportunity for a data collection method to have an impact on the reported data.

One of the major results of the study is that at the seven state level the only univariate test which showed a difference among the methods was the test on nonresponse rates. The personal interview method had a lower nonresponse rate than the other three methods of data collection. Although differences in nonresponse rates were not strictly significant, the significance level of the data, 0.11, was low enough to make this finding important. Experience indicates it is easier for many farmers simply to ignore questionnaires that they receive in the mail or hang up on telephone enumerators than to refuse a field enumerator standing in front of them.

On the seven state level, the four hog variables did not show any significant effect due to the method of data collection when univariate tests were run. For some variables the sample size was not large enough to determine some rather large differences as significant. For example, differences of twelve and thirteen percent for variables 3 and 4 were not significant. Thus, the power in these tests was probably not acceptable.

Although no univariate tests were significant, a multivariate test on the positive data showed that the effects of the methods across all four variables was significant. Indeed, the significance level was much less than 1 percent (see Table 4) for estimates at a seven state level. The reason for this significance can be seen in Table 3. The personal interview method tended to have the opposite effect of the mail-personal delivery method on the means of variables 1, 3, and 4. (The means for variable 2 were almost equal.) The advantage of a multivariate test over a set of univariate tests is the ability to take into account relationships across all variables. While the personal interview method gave the highest estimate of variable 1 and low estimates of variables 3 and 4, the mail-personal method gave the lowest estimate of variable 1 and high estimates of variables 3 and 4.

The difference in these two procedures might be attribute to the fact that the mail-personal delivery method tends to minimize the effect of the enumerator while the personal interview method tends to maximize the effect of the enumerator. Is this enumerator effect good or bad? The enumerator effect is good if the enumerator explains terms and questions to the respondent so that the questionnaire is completed accurately. The amount of editing required on mail questionnaires had indicated that the respondents do have difficulty in completing the questionnaire without enumerator assistance. The enumerator effect is bad if the enumerator's presence, attitude, appearance, etc. cause respondents to bias their answers. Because testing of enumerator effects for personal interviews is usually difficult and expensive, the U.S. Department of Agriculture has not studied this effect. This study shows that such an investigation is warranted.

A significant interaction exists between the states and the data collection methods at both a univariate and multivariate level. For example, when a multivariate test was run on the inter-

TABLE 3

Means for Four Hog Variables and Nonresponse Rates for Four Methods of Data Collection*

| Data Collection Method | All Data | | | | | Positive Data | | | |
|--|------------|------------|------------|------------|------------------|---------------|------------|------------|------------|
| | Variable 1 | Variable 2 | Variable 3 | Variable 4 | Nonresponse rate | Variable 1 | Variable 2 | Variable 3 | Variable 4 |
| Operational (Mail-Telephone-Interview) | 122.6 | 6.8 | 8.0 | 6.4 | 0.13 | 304.3 | 16.9 | 19.5 | 16.5 |
| Mail--Personal Delivery | 119.4 | 6.9 | 8.0 | 6.6 | 0.14 | 286.1 | 16.8 | 19.3 | 17.5 |
| Telephone--Personal Interview | 123.2 | 6.9 | 8.5 | 5.8 | 0.12 | 299.1 | 16.9 | 20.2 | 15.1 |
| Personal Interview | 126.4 | 7.0 | 7.5 | 6.3 | 0.09 | 312.5 | 16.8 | 17.8 | 15.7 |

| | | | | | | | | | |
|-------------|-----------------|---------------|---------------|---------------|----------------|-----------------|----------------|----------------|----------------|
| All Methods | 122.9 (0.78) | 6.9 (0.98) | 8.0 (0.31) | 6.3 (0.53) | 0.12 (0.11) | 290.9 (0.21) | 16.8 (0.96) | 19.2 (0.21) | 16.2 (0.16) |
|-------------|-----------------|---------------|---------------|---------------|----------------|-----------------|----------------|----------------|----------------|

*Significance levels of the data when testing the hypothesis that the four methods of data collection yield different estimates are in parenthesis in the bottom row. Significant differences between the four interviewing methods is indicated by a value in parenthesis which is less than or equal to 0.10.

actions, the significance level of the data was .005 for all data and .001 for positive data. A significant interaction indicates that relationships at the seven state level are simply cumulative effects and that relationships of the methods are contradictory from one state to another. For example, in Indiana the estimates for positive data have almost the opposite relationship of those in Iowa. Therefore, the conclusion is that the method of data collection is important in conjunction with the conditions that exist in the state where each method is applied. Office procedures, local conditions that affect respondents' attitudes, and individual enumerators are examples of conditions that might contribute to the state effect. Because of this interaction, it would be incorrect to study the differences among the data collection methods in one state and infer that similar differences exist in all states.

At the state level, the univariate tests showed the method of data collection did have a significant impact on the nonresponse rate in Wisconsin. The nonresponse rates in Wisconsin were 4 percent for the personal interview and telephone-personal interview methods and 10 percent for the operational and mail-personal delivery methods. The significance of the nonresponse rates in Wisconsin is important because Wisconsin is the only state where operators in the sample were not previously contacted during previous quarters. This fact implies (but does not prove) that the previous survey experience of operators in the other six states tended to nullify the effects of the data collection methods on nonresponse rates. Thus, the effect on nonresponse rates in seven states with new samples may be greater than the effect on the seven states used in this study.

Results When Outliers Are Omitted

A very larger report (outlier) in a group may by itself have a large effect on the mean of the group. There are two ways to view the impact of these large observations. One could say that a few outliers will cause any method to which they are assigned to be significantly different from the other methods. One the other hand, there is the possibility that the specific method may have caused the outliers. For this study the first viewpoint seems much more likely than the second.

TABLE 4
Significance Levels of the Data for Wilk's
Multivariate Test of Differences
Among Four Methods of Data Collection

| State | Multivariate Test (Wilk's Statistics) | |
|-------------------|---------------------------------------|--------------------------|
| | All Data α value | Positive Data α value |
| Illinois | 0.79 | 0.41 |
| Indiana | 0.10* | 0.55 |
| Iowa | 0.11 | 0.01* |
| Minnesota | 0.12 | 0.13 |
| Missouri | 0.94 | 0.47 |
| North Carolina | 0.85 | 0.03* |
| Wisconsin | 0.76 | 0.61 |
| 7 States Combined | 0.21 | 0.0007* |

This test is only on four survey variables and does not include the non-response rate.

*Significant differences among the four methods are indicated by a value $\leq .10$.

Estimates with the outliers removed did not differ much from the analysis that included the outliers. Differences among the methods for positive data are still not significant at the seven state level for any of the four survey variables. Also, the multivariate test on the positive data was still significant.

TABLE 5
Significance Level of the Data for the Hypothesis that the
Four Data Collection Methods Yield Equal Estimates
(Outlying Data Points Removed)

| State | Positive Data | | | |
|----------------|---------------|------------|------------|------------|
| | Variable 1 | Variable 2 | Variable 3 | Variable 4 |
| Illinois | .98 | .85 | .87 | .35 |
| Indiana | .48 | .62 | .89 | .31 |
| Iowa | .03* | .78 | .21 | .47 |
| Minnesota | .81 | .77 | .24 | .72 |
| Missouri | .53 | .42 | .59 | .50 |
| North Carolina | .71 | .71 | .19 | .18 |
| Wisconsin | .12 | .52 | .35 | .17 |

| | | | | |
|-------------------|-----|-----|-----|-----|
| 7 States Combined | .21 | .94 | .16 | .14 |
|-------------------|-----|-----|-----|-----|

*Significant differences among methods are indicated by an α value $\leq .10$.

SUMMARY

Three major findings resulted from this study:

1. On a seven state level the personal interview method yields a lower nonresponse rate than the other three methods of data collection. The lower nonresponse rate is important because nonresponse can affect all survey estimates if the nonrespondents have a different distribution from the respondents.
2. Although the data collection methods have no significant impact on survey estimates for univariate tests at a seven state level, a multivariate test on only those respondents who have had a positive number of hogs was highly significant. This significance was due to differences in the personal interview method and the mail-personal delivery method. The differences in mail-personal delivery and the personal interview method probably deserve further investigation. Despite the fact that neither of these two methods are currently used for list surveys in the U.S. Department of Agriculture, mail and personal interview are part of the operational procedures of others surveys in the Department.
3. A significant interaction exists among the states and the various data collection methods (at both a univariate and multivariate level). Consequently, local conditions in individual states may cause the estimates from data collection methods to exhibit different relationships from one state to another. In this study the differences from state to

state tended to balance out over all seven states and resulted in no significance for univariate tests on the survey estimates.

Besides comparing the data collection methods, this study also illustrates the need for several states in many research projects at the U.S. Department of Agriculture. Not only may one state yield estimates with too much variability to give useful results, but also one state may have local conditions which cause results contradictory to the results from other states. In this study the highly significant interaction between states and data collection methods gives firm evidence of contradictory results in different states.

All results and conclusions in this study are made with the knowledge that the sample size was not large enough to detect any significant differences less than 10 percent. Of course, differences less than 10 percent may be important, but using more than seven states for a research project puts an obvious strain on any operational survey which is concurrent with the research project. Although the U.S. Department of Agriculture recognizes this strain, it must also recognize that future research projects may sometimes require seven or more states in order to be worthwhile projects.

REFERENCES

1. Barr, Anthony J., James H. Goodnight, John P. Small and Jane T. Helwig, A User's Guide to SAS 76, SAS Institute, Raleigh, North Carolina. 1976.
2. Steel, Robert and James Torrie, Principles and Procedures of Statistics, McGraw Hill Book Publishing Co., Inc., New York. 1960
3. Timm, Neil H., Multivariate Analysis with Applications in Educational Psychology, Brooks/Cole Publishing Co., Monterey, California. 1975

APPENDIX

1. Weighting the data:
Weights were assigned to each variable to reflect the relative expansion factors. Weights indicated relative importance of each stratum in each state across all seven states.
2. Replication:
The data set was ordered by state, data collection method, stratum, crop reporting district, county and an identification number. In this ordering the data set was systematically divided into ten replicates for analysis purposes. Mean values for each treatment were calculated within each replicate, and combined in the following manner:

$$\bar{x}_{sr} = \text{sample estimate of the weighted mean in the } r^{\text{th}} \text{ replicates, } r = 1, 2, \dots, 10, \text{ and in the } s^{\text{th}} \text{ state, state} = 1, 2, \dots, 7$$

$$\bar{x}_s = \text{samples estimate of the weighted mean in state } s$$

$$\bar{x}_r = \frac{\sum_{s=1}^{10} \bar{x}_{sr}}{10}$$

$$\bar{x}_r = \text{sample estimate of the weighted mean of the } r^{\text{th}} \text{ replicate over all states}$$

$$\bar{x}_r = \frac{\sum_{s=1}^7 \bar{x}_{sr}}{7}$$

\bar{x} = sample estimate of the weighted mean over all states and strata

$$\bar{x} = \frac{\sum_{s=1}^7 \bar{x}_s}{7} = \frac{\sum_{r=1}^{10} \bar{x}_r}{10}$$

The unbiased estimate of the standard error for each state estimate is:

$$SE(\bar{x}_s) = \frac{\sum_{r=1}^{10} (\bar{x}_{sr} - \bar{x}_s)^2}{9(10)} \quad 1/2$$

An unbiased estimate of the standard error for an estimate of a seven state total is:

$$SE(\bar{x}) = \left[\frac{\sum_{r=1}^{10} (\bar{x}_r - \bar{x})^2}{9(10)} \right]^{1/2}$$

Besides simplifying the calculation of standard errors, the use of replicate values in the statistical analysis:

- 1: assured equal cell sizes in the analysis of variance
- 2: yielded distributions which are fairly normal.

These benefits of replication make the analytical exploration of data from a complex survey design much more straightforward and accurate. Univariate and multivariate tests were conducted using the SAS computer package. The processes involved in this analysis are outlined below, although more details are available in the SAS 76 User's Guide.¹

A general linear model was constructed to fit the values from the replications: $\underline{Y} = \underline{X} \underline{B} + \underline{E}$, where \underline{E} is the residual error matrix, \underline{Y} is the vector of estimated means for four (quantitative) hog variables, \underline{X} is

the data matrix, and \underline{B} is the vector of model parameters. The model parameters consist of three effects: the state, the treatment, and the state-treatment interaction. The treatment effect refers to the method of data collection discussed in the background section of this paper. The MANOVA option of the GLM procedure in SAS was used to test the hypothesis that 1) there is no difference in the data among treatments and 2) there is no state-treatment interaction. Both univariate and multivariate tests were run on the four survey variables, and an univariate test was run on the response rate. Duncan's new multiple range comparison test was used on those variables for which the various treatments proved to be significantly different. The test was used to determine which, if any, of the individual treatments yielded values that were significantly different from the other treatment means. Details of the process may be found in Principles and Procedures of Statistics, by Robert Steel and James Torrie², and the SAS

76 Guide. Basically, the procedure ranks the means of the various treatments in increasing order. A difference is declared significant if its absolute value exceeds the appropriate test value, where the test value is determined by:

$$T \cdot S \frac{1}{2} \left(\frac{1}{t_i} + \frac{1}{t_j} \right), \quad i = j$$

where t_i = the number of observations in the i^{th} treatment group,

$$S = \frac{\text{Mean square error}}{\text{Number of observations per treatment}}$$

T = the value from the appropriate Duncan's test table.

Wilk's Λ criterion, which was used for the multivariate tests, is essentially a multivariate extension of the F test used in univariate analysis of variance. Full details on Wilk's Λ criterion are in Timm's book.