THE NASS RESPONSE TO NONRESPONSE

by Diane K. Willimack

Mother Goose surveyed the population of pigs in nursery rhymes to determine which ones "had roast beef." The results were:

SAMPLE UNITS	RESULT OF CONTACT	RESPONSE CATEGORY
"This little pig went to market"	not found at home	inaccessible
"This little pig stayed home"	slammed the door on Mother Goose and would not answer	refusal
"This little pig had roast beef"	obtained answer to question	respondent
"This little pig had none"	obtained answer to question	respondent
"This little pig cried, *whee, whee, whee,* all the way home"	answered Mother Goose with "I don't remember"	unknowing
the pig stolen by Tom, the piper's son	Mother Goose forgot this pig was even part of the population	not covered

Like the above survey of nursery rhyme pigs, any NASS survey suffers the potential for nonresponse from these sources: (1) REFUSAL, where the respondent will not answer the questions; (2) UNINFORMED or UNKNOWING of the answer to the questions being asked - the "don't knows"; (3) INACCESSIBLE, those potential respondents that cannot be found during the period of data collection; and (4) NONCOVERAGE, units that should have been included in the sample but were incorrectly excluded.

Inaccessibles and noncoverage result in entire units being lost to the sample. Refusals and "don't knows" may also result in the loss of entire units to the sample, but they may occur for individual items within a questionnaire as well. For example, an operator may refuse to answer questions about expenses, or the spouse of an operator may not know the amount of grain stocks during an Crops/Stocks Survey interview.

Regardless of its source, nonresponse is a nonsampling error that can seriously undermine the accuracy of survey estimates. Accuracy refers to the difference between what we actually measured and what we wanted to measure. This difference is called the bias in the estimate.

Why can nonresponse be so damaging? If the original sample was random, aren't the remaining respondents still a random sample? The answer is yes only if the nonrespondents occurred randomly as well. However, this is rarely the case. For example, in A Sampler on Sampling, one of the references for this series, the author describes "the case of the missing children," where the ratio of children to adults in a neighborhood appeared to decline over time. This situation occurred because, in the initial contact by interviewers, adults with children were more easily found at home in the daytime than were adults without children. As the survey was repeated, more contacts were made in the evening, and the adults without children were found. The nonrespondents (inaccessibles) to the initial contact were quite different from the respondents.

Examples of nonresponse bias can also be found in NASS surveys. In a research study in 1978, Chap Gleason and Ron Bosecker found that nonrespondents to cattle or hog surveys tended to have more of these animals than did the respondents and that larger operations tended to refuse more than smaller operations. Livestock indications, at that time, tended to be too low. Also, in research by Dave Dillard in the late 1980's, characteristics of farms captured by new screening procedures in the annual economic survey are compared to those captured by the old procedures. These data tell us how different these potential nonrespondents (due to noncoverage) are from the rest of the respondents.

The amount of damage done to an estimate by nonresponse depends on two things: (1) the number of the nonrespondents relative to the total sample size and (2) the magnitude of the difference between the nonrespondents and the respondents. For instance, consider farm operators who refuse the economic survey or are coded inaccessible because they have extended their tax filing date. If these nonrespondents are a very small proportion of the total sample size, say only 1 percent, their financial information will have little effect on the estimate, even if their expenses or incomes are quite different from the respondents. On the other hand, if there is a large proportion who delay tax filing, but their expenses and income are similar on average to the respondents, their effect on the estimates will again be negligible. However, if the delayed filers account for a large proportion of the total sample size, say 20 percent, and if they tend to be large operations, the data from these nonrespondents would significantly affect the estimates.

The previous examples show the need for efforts to minimize nonresponse. In NASS, such efforts include public relations that publicize the purpose of our surveys, offers of incentives (such as copies of published releases), adequate survey training so that the survey is conducted efficiently, sustained courteous relations between enumerators and respondents, and telephone follow-up after mailing, or personal contact after telephoning. Nevertheless, nonresponse occurs. Fortunately, there are methods to adjust an estimate for nonresponse. In some cases, NASS estimates are "reweighted" by adjusting the expansion factors to reflect the nonresponse rates. This procedure assumes that the nonrespondents within each stratum are like the respondents. Another procedure is to draw a sample of the nonrespondents and make a special effort to obtain the required information from them. Then nonresponse bias can be

measured, and the overall estimate can be adjusted accordingly.

When only a few items are missing from an entire questionnaire, it may be possible to impute, or estimate, these data from other available information. In NASS, for example, during the questionnaire edit phase, statisticians in a Field Office are often called upon to perform “deductive imputation," where the statistician attempts to deduce the missing values based on logical relationships with other data collected, enumerator notes, previous data collections on the same unit, or control data on the list frame. Other common imputation procedures include the substitution of the overall mean or the stratum mean for the missing value. This approach is used by NASS in estimating grain stocks. Finding out whether an operation has stocks is valuable here because imputation need not occur for those with valid zeroes. That is, we do not want to impute for missing values more often than we should.

Many other imputation schemes are available. A value for a missing item may be taken from another respondent in the sample, selected either in sequence or at random. Respondents may be matched on certain relevant characteristics before a value is borrowed. An equation or formula could be used to calculate the missing item from related items or from control data. This last approach is often used for the automated imputation of the Agricultural Surveys.

Nonresponse has the potential of nullifying the most carefully determined sample size, playing havoc with the most appropriate sample design, and destroying the accuracy of the most precise estimate. Yet all is not lost. Though obtaining responses is by far the best defense against the perils of nonresponse, the integrity of NASS estimates is maintained by proper and appropriate techniques for studying the characteristics of nonrespondents, measuring nonresponse bias, and adjusting indications accordingly.