United States
Department of
Agriculture

National
Agricultural
Statistics
Service

Research and
Applications
Division

NASS Research Report
Number SRB-88-01

January 1988

The Influence of Using Previous Survey Data In the 1986 April ISP Grain Stocks Survey

Bradley V. Pafford

THE INFLUENCE OF USING PREVIOUS SURVEY DATA IN THE 1988 APRIL ISP GRAIN STOCKS SURVEY, by Bradley V. Pafford. Research and Applications Division, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, DC 20250. January 1988. NASS Research Report No. SRB 88-01.

ABSTRACT

Proper use of previously collected grain stocks data in the current grain stocks interview was studied in a Computer Assisted Telephone Interviewing (CATI) environment. This effect along with effects due to the size of the farm operation, CATI telephone interviewer, and respondent were analyzed with data collected from the April 1986 Integrated Survey Program (ISP) survey in three States. The direct use of prior grain stocks resulted in higher grain stock levels in the April responses especially for large farm operations. Also, there were more instances of reporting problems for stocks and storage capacities when the respondent changed between the quarterly surveys. The reporting problems include responses where the accuracy of the January report was questioned by the respondent or a comment was given indicating incorrect reporting in either survey.

KEYWORDS: Historical data, analysis of variance, tests of independence, and grain stocks

ACKNOWLEDGMENTS

The author would like to thank the many people who helped plan and implement the study and analyze the results. Thanks go to Gene Danekas and Dick Coulter from the CATI staff. Thanks also go to Doug Kleweno who helped plan the study and Gretchen McClung who helped with the analysis. Special thanks go to the staff of the California, Georgia, and Nebraska field offices who cooperated tremendously under otherwise heavy workloads.

CONTENTS

SUMMARY

A research study was undertaken for the April 1986 Integrated Survey Program (ISP) Survey (now called the Quarterly Agricultural Survey Program) in California, Georgia, and Nebraska to study the effect of previously reported grain stocks data in a Computer Assisted Telephone Interviewing (CATI) environment. The effect of providing January ISP grain stocks data in the grain stocks portion of the April survey was measured through split sample testing procedures. A control group (no use of prior data, which is the operational survey procedure) was compared against an experimental group (direct use of prior data). Analysis simultaneously evaluated the historical data treatment effect with effects due to the interviewer and size of the farm operation. Reasons for changes in stocks and capacity were evaluated by analyzing data from CATI probing questions, and from enumerator post-survey comments. The effects of changing respondent between quarterly surveys were studied.

It was found that the experimental group's stocks estimates were significantly larger than those for the control group for some crops, and that this supported the research hypothesis. This hypothesis was that the experimental group would more than likely report closer to their January response than the control group, and this should produce a higher response since the January 1 stocks levels are, on average, larger than in April. Further analyses revealed that these differences occurred mostly in one stratum, which represented large farm operations.

Also, interviewers, on average, obtained the same mean farmer responses, and this relationship was the same no matter whether they interviewed with the control or experimental group samples. In essence, there were no consistent effects due to the interviewer.

Reasons for corn storage increases from January to April were investigated. When this occurs during a survey, the respondent often needs to be re-called to resolve the apparent discrepancy. It was found in a large percentage of cases that when corn storage increased these may not have been real increases. That is, problems in getting the correct January corn stocks were noted from answers to standard CATI probes. The same relationship existed for changes in storage capacity. For example, in over three-fourths of the cases when storage capacity changed from January to April, the reasons for these changes could be classified as "problematic." Problematic responses were ones where the accuracy of the January report was questioned by the respondent, or a comment was given indicating incorrect reporting in either survey.

Responses classified as problematic were compared with changes in respondents between surveys. Significantly more problematic responses were found when the respondent changed between the two quarterly surveys, compared with when the same respondent reported. Finally, CATI office experience played a role in the reaction of the enumerators to use of historical data in the interview process.

It is recommended that prior grain stocks data not be used directly in the current grain stocks interview, and that research focus on use of these data as an editing tool to be used after an initial response is obtained. In addition, it is recommended that more use be made of prior information interview, such as who the respondent was in the previous quarter's survey.

BACKGROUND

The National Agricultural Statistics Service (NASS) is an agency of the U.S. Department of Agriculture responsible for collecting and publishing a wide range of agricultural statistics on a weekly, monthly, quarterly, and yearly basis. The quality of these published data has always been a primary concern. NASS makes extensive use of previously collected data for improvement of current collected data. The current response is traditionally compared with some historical response after the interview has taken place. If responses differ considerably, or are outside a pre-specified range, the respondent is often called back to resolve the discrepancy. There is interest in NASS to use previously reported data during the personal or telephone interview, or provided on mail questionnaires. It is this use that concerns NASS policymakers because of its potential to bias the respondents answers. Little research has been published or found in the literature to provide direction. Beginning in 1985, the NASS Research Division decided to formally evaluate proper uses of historical data.

Integration of some individual commodity surveys into an "Integrated Survey Program" (ISP) has been the focus of the NASS multiple frame survey program since at least 1984. This effort involves organizing the list portion of these prior individual surveys through extensive stratification of NASS's list frame of farmers (special use strata are often constructed), and through use of replicated sampling procedures. One of the outcomes of the ISP is that there is a greater need and interest in accessing historical responses. This has been accentuated by the recent availability of computers that allow for efficient data storage and retrieval, Computer Assisted Telephone Interviewing (CATI), and high-speed laser printers.

Pafford (1986) began the study of proper use of historical data in a limited framework for NASS surveys (one State, survey and year using CATI). In that study, crop planted acreage questions were asked that also appeared in another survey 5 months earlier. Responses were expected to be the same since the crops were planted before the first survey. Sample units were randomly assigned to one of three "historical data treatment" groups and responses compared. The treatments ranged from no use of previously reported planted acreage to direct use of these data in the current interview. The results indicated that farmer responses were affected in proportion to the level of use of these prior survey data. For example, a significant difference was found between the acreage reports when no direct use was made of the response from the first survey. However, no difference could be uncovered when the respondents were directly given their earlier responses during the CATI interview. The next question was, "How might responses be affected in other types of NASS surveys, such as grain stocks surveys?"

As a continuation of that research, this paper presents findings from a project developed in conjunction with the 1986 April ISP Survey in Nebraska, California, and Georgia. The project's purpose was to determine what effect providing previously reported grain storage levels (collected 3 months earlier) would have on the current report of grains on hand. Responses were expected to change from one survey to the next. That is, grain stock levels are not fixed, but grains are regularly bought, sold, used for feed on livestock farms, and moved in or out of government programs. Other factors which could affect responses or explain discrepancies were studied, such as the interviewer, type of respondent, and size of the farm operation. In this study, differences by respondent relate to comparing two groups: same versus different respondents in consecutive quarterly ISP surveys. Interviews were completed using the University of California at Berkeley Computer Assisted Survey Execution System (CASES).

The objectives of this research project were:

1. To compare responses to grain stocks and storage capacity questions when no use is made of previous stocks information, versus direct reference made to these data during the current interview.

2. To evaluate other factors affecting responses, namely the interviewer, respondent, and size of operation.

3. To document reasons respondents gave for changes in the stocks levels between January and April in the case where historical data were used in the interview.

4. To summarize interviewer perceptions of use of historical data as they were obtained from a enumerator post-survey questionnaire.

LITERATURE REVIEW

The only formal research done in NASS relating to use of historical data was by Pafford (1986). Here, previously reported planted acreage was provided in different controlled "treatments" to measure the effect on the current report of this same planted acreage. The current survey was the 1986 Fall Acreage and Production (A&P) Survey in California, with reference back to the 1986 June Acreage Survey. Planted acreage was asked in both surveys. Responses were expected to be the same, as the acreage in question was planted prior to the June Acreage Survey. Three ways, or "treatments," to provide previous planted acreage were evaluated using a highly structured CATI interviewing environment.

The results were the following. Re-reporting of June planted acreage was much more likely to occur when the interviewer provided the previous response (June planted acreage) directly to the respondent (TREATMENT 3), compared with when no reference was made to these data (TREATMENT 1). In other words, there were small mean differences in the Fall and June responses when the historical data (June planted acreage) was directly referenced in the interview. The largest mean difference between the June and fall responses were noted when no reference was made to these data. If these historical data were available in the margins of the CATI screen (TREATMENT 2), with instructions to the enumerator to never directly reference them, then mean differences fell in between TREATMENT 1 and 3 levels.

Lack of information on the "true" planted acreage limited interpretation of the data. However, if the June response was considered correct, then providing historical data tended to improve the accuracy of the Fall A&P Survey. Intuition would lead one to believe the June response was closer to the true value since actual planting took place nearer this date than the following fall in California.

Another interesting finding was that by providing historical data directly in the interview, the respondent was less likely to "forget" (not report the survey item in one of the surveys) to report the acreage in question. Also, less "forgetting" was noted for the reporting of the row crop planted acreage than the small grain acreage. The reason speculated for this was it was easier to recall plantings 6 months earlier (row crops are planted in the spring) than plantings of a year ago (small grains are planted in the fall).

Analyses of enumerator effects, size of operation effects, and interactions of the

enumerator, size of operation (strata), and treatment were possible due to the study design (that is, factorial experiment). No important interactions were noted. That is, for example, the treatment response was the same for each type of operation, and the enumerator response was the same for each treatment level. Significant differences were found between the strata levels (that is, stratification was effective). Finally, mean responses across enumerators were not statistically different; there was no enumerator effect.

While these results were of value, they were limited in scope to just one State, one time period, and one survey. Would the tendency to report the same figures exist in a different setting? In NASS surveys, the same questions are typically asked in consecutive surveys, but responses are expected to change. Individual farmer grain stock levels can change dramatically between quarters as some can sell, buy, or feed large quantities of corn or soybeans in a few days. What effect would use of previous grain stocks data have on the current response to grain stocks questions? Also, other factors could affect the interview, such as the change in respondent. For example, consistent reporting should be expected if the farm operator reports in both time periods. Less consistent reporting should exist when the farmer's report is followed by the spouse's report. How will the introduction of historical data affect this proposed relationship? Finally, other information may prove helpful in the interview that were not used in the study by Pafford (1986). Can we demonstrate that introducing who reported the historical data and the date of that report will not be objectionable to the respondent, and perhaps will be of value?

Literature found in journals, for the most part, does not address the issue of use of historical data (see Pafford, 1986) for a complete review of related literature). An overview of this research is now given.

First, the primary focus in the literature is on response variability, such as simple and correlated response variance (Bailar, 1968, Kish, 1962, Hansen, Hurwitz, and Pritzker, 1964). Simple response variance (SRV) is trial-to-trial variability. Conditions for measuring SRV are that initial and followup interviews are conducted, the survey conditions are as similar as possible, and the questions reference the same data item. No reference to the earlier response is generally allowed in these type of studies. In only a few papers is the issue of use of previous survey data considered. O'Muircheartaigh (1986) found lower estimates of SRV when historical data (initial interview data) were accessible by the reinterviewer for use in problem resolution (not to be used until after an independent reinterview response was obtained) compared with when these prior data were not accessible during the reinterview. That is, reinterview responses tended to be more consistent with the original responses when the interviewer had this knowledge, than when no knowledge of the previous responses was available. Interviewers, then, were reacting differently to various survey conditions (such as reinterviewing husbands when the wives were first interviewed) when they had knowledge of the initial response. O'Muircheartaigh concluded that instructions to enumerators on use of initial interview data, which direct them not to use it until after the reinterview response has been obtained, were not being followed. Use of initial interview data in the Current Population Survey (CPS) was questioned. Hansen, Hurwitz, and Pritzker (1964) first noted this type of response effect in their early evaluations of the CPS reinterview program. The interviews were conducted through personal interviewing. This allows more flexibility in inappropriate access to the previous response when compared with the more structured central telephone interviewing environment, especially CATI environments.

In other CPS related research, Bailar (1968) studied the length of time between a

reinterview and enumerator access to the respondent's original response. The results indicated that the best procedure was where the reinterview was done as soon as possible after the initial interview, and where the interviewers did not have access to the original responses.

Studies on interviewer variability are numerous in published journals. Relatively few studies exist in NASS in this area. Interviewer variability is the component of response variance reflecting the propensity of interviewers to "do things different" from each other. Early work for personal interviewing situations was done by Mahalanobis (1946), Hansen and Marks (1958), Hansen, Hurwitz, Marks, and Maudlin (1951), and others. The largest interviewer effects were noted when some resistance on the part of the enumerator to the question existed, when questions were ambiguous and wordy, when additional probing was needed, and when attitudinal questions were needed versus more factual questions such as number of corn acres planted on a farm.

Work on interviewer variability in the telephone environment came later. Groves and Kahn (1979) showed that telephone interviewer variability was generally less than for personal interviewing. However, the effect was greater since there are more interviews done by a telephone interviewer. Groves, Magilavy, and Mathiowetz (1981) compared highly controlled (monitoring) with less controlled (no monitoring) telephone environments. General results from these and other studies are similar to those for personal interviewing. Additionally, low interviewer variability can be achieved by close supervision, monitoring of performance, and exchange of interviewing techniques among interviewers. Probing and other uncontrolled interviewer procedures tended to increase variability.

The effects due to the respondent are also important. NASS has long noted these differences in the farm population they sample. Steiner (1980), for example, reported many more corrections in previously reported figures when the respondent changed between two consecutive interviews. Bosecker (1977) reported different levels of total acres, tract acres (acres in a tract of land), and total cattle by respondent group. Warde (1986) reported smaller acreage and counts of hogs and pigs for spouses than for farm operators. Finally, Nealon and Dillard (1984), in a nationwide study of characteristics of farm wives and husbands, found significant differences in the reporting of some characteristics. Wives had significantly more missing data and lower mean responses than husbands for an overwhelming number of characteristics.

In journal literature, O'Muircheartaigh (1986) found SRV lowest for self-self reports. Less reliable reporting occurred for the proxy-(same) proxy group, with the least reliable group being the proxy-(different) proxy respondent group. Respondent characteristics, such as age, husband/wife, or relationship to head of household also affected consistency of reporting.

STUDY DESIGN

Historical Data Treatment Procedures

This study evaluated two historical data "treatment" procedures. The historical data were the stocks and storage capacity reported for the 1986 January ISP survey. These data were not referenced in the 1986 April ISP survey for one-half the sample units (the control group) and were referenced for the other half of the sample units (the experimental group). For the control group, then, no stocks or capacity information

was made available to either the interviewer or the respondent. These responses were compared against the experimental group which were given their January stocks and capacity responses immediately prior to asking the corresponding question. In this way an upper bound or maximum effect of providing previous stocks data directly in the interview could be studied. More traditional and envisioned uses of these data have been as an on-line editing tool that allows the respondent access to the previous response if the current response deviates appreciably. However, it is ideal to assign treatments to sample units prior to the interview for split-sample testing purposes. In this way the outcome of the response does not influence the assignment of the treatment procedure.

Examples of the wording of the questions can be found in Appendix A. More discussion follows in the section on Data Collection Methodology.

Sampling Design

The study had to be designed around the NASS list survey program sampling plan which is stratified simple random sampling with replicated samples in each stratum. Sampling specifications were for a 50-percent overlap in sample units between consecutive quarterly surveys (drawn easily by adding and deleting replicates). Within this framework, replicates that were sampled in January were assigned to the experimental group, and replicates new to the April Survey were assigned to the control group. This created a split-sample testing situation with an approximately equal number of cases in each treatment.

A more involved design was actually used to allow for simultaneous comparisons of treatment, strata, interviewer, and their interactions. A two-factor factorial design in blocks was set up with factors, enumerator and historical data treatment procedure, and blocks the sampling strata. Three strata levels were used representing farm operations with "small," "medium," and "large" acreage. Sampling units were randomly assigned to each stratum by treatment by enumerator combination. Every enumerator received every treatment by strata combination. A near balanced designed was purposely achieved to avoid complexities during the analysis. To assist another research project concerned with optimal times to call farm operators, an additional randomization was incorporated. Specifically, three time of day calling "slots" were set up for first calls only with time periods: 3:00 to 5:00 p.m, 5:00 to 7:00 p.m., and 7:00 to 9:00 p.m. This effect is ignored for the purposes of this study. One can read Warde (1987) for these results.

Data Collection Methodology

In this study, one interviewer instrument was used for both treatments. The treatments were randomly assigned (as previously discussed) and a treatment number was included as part of the case identifier within CASES (case number appeared on the call sheet). By keying on this number, appropriate branching within the instrument controlled which set of questions (treatment) was used for a particular respondent. Telephone interviewers did not know prior to the interview which treatment was being used.

Three folders were given for each enumerator, one each for each time slot (see previous section). Each folder contained "call sheets," or pieces of paper that had the CATI case number to call up on the computer, and spaces for recording the outcome of

each call. After the initial call, if the case was completed, the case went into the completed cases. If the case had to be recalled, a folder was provided for callbacks. Callback appointments were handled individually. Enumerators switched to the next folder for the next time period following the same routine for handling completed cases and callbacks. After all cases had been called once, the time folders were discarded. Little inconvenience was noted. The randomization of enumerator and treatment assignments was handled by the author prior to being sent to the field. For example, each folder contained, as near as possible, an equal number of cases in each treatment group and strata. Enumerators also received every combination of strata by treatment by time slot, with as near as possible an equal number of observations. To insure that certain assignments were not being favored, the professional staff or telephone supervisor was instructed to nightly randomize each enumerator's call sheets.

Data files from the previous (January) grain stocks survey were accessed and selected items prestored in the data files for use in the interview instrument for the current survey. These selected items were: the January stocks and storage capacity report, the date of the January response, and a code for the January respondent (operator, spouse, other). The previous response and the date of that response were incorporated into the wording of the experimental group's questions. The previous respondent was displayed on the screen for interviewers to use if a question about who supplied the January response was raised. Otherwise, it was not used (see Appendix A).

The actual wording for the two treatments for stocks was as follows.[1] Corn is used for illustration as are a January 3 previous date and a 2,000-bushel previous response.

Control Group: How many bushels of CORN are stored on your operation ?

Experimental Group: Our records show that on January 3 your operation reported 2000 bushels of CORN being stored.

How many bushels of CORN are now being stored ?

For grain storage capacity, the analogous wording was:

Control Group: What is the CAPACITY of GRAIN STORAGE facilities located on the total acres you operate ?

Experimental Group: Our records show that on January 3 your operation reported to us a GRAIN STORAGE CAPACITY of 15,000 bushels.

Is this still correct OR has your storage capacity changed since this report?

For the experimental group, if the absolute value of the difference between the current reported stocks and the previously reported stocks exceeded a predetermined value (100 bushels was chosen - see footnote 4), then the next screen for the interviewer was a probing question. The question was either "Did you buy any CORN ?" or "Did you sell any CORN ?" depending upon whether the current report was greater than or less than

[1] NASS reference dates have since been changed to the first of the month for grain stocks reporting.

the previous report. If the answer to this question was "no" then explanatory notes were solicited from the interviewer, such as was it used for feed, or included in government programs.

By using CASES, all control over assigning treatments to sample units was removed from the interviewer. Also, since branching was controlled within the instrument, a minimum of additional training was needed beyond what was necessary for an operational CASES survey. All interviewers had some experience with CASES, but for some, particularly in Georgia, this experience was limited to one or two small surveys. In all cases, this was the first experience with the grain stocks questions on CASES. One training session of about 4 hours, primarily devoted to mock interview practice, preceded actual interviewing.

ANALYSIS PLAN

Analyses of these data proceeded as follows. First, these data were modeled using analysis of variance techniques. The factors of size of operation, historical data treatment procedure, and interviewer were considered together to see which contributed, either individually or in combination with the other factors, most to the overall survey variability. See Appendix B for a formal presentation of the model, the weighting procedures employed, and the procedure for constructing appropriate tests of hypotheses.

Second, further analyses were done to compare the estimated State total grain stock levels by historical data treatment group. Third, analyses were done to explain discrepancies in the data. These were in the form of simple frequency distributions, and, in some cases, tests of independence for 2x2 contingency tables (see Appendix C for a discussion of chi-square tests for independence in sample survey situations). Finally, enumerator comments were summarized.

Table 1 shows the number of samples drawn in each State and the response distribution.

Table 1. Response distribution by State--1986 April ISP Survey 1/

State	Sample size	Inaccessible	Out of business; refusals	Usable responses	Percentage usable
			Number		Percent
California	642	126	45	471	73
Georgia	1,264	260	158	846	66
Nebraska	1,611	209	296	1,106	68

1/ Percentage usable reflects response via CASES. Personal interviews were used to follow up for nonresponse.

RESULTS

Modeling

Tables 2 and 3 present the results of model tests for Georgia and Nebraska. One model was built for each crop (corn, wheat, and soybeans stocks) and onfarm storage capacity in each State. California results were not presented because little crop storage of any kind existed for the farmers in the sample.

The effects considered in these two tables, listed under "Sources of Error" were that due to the size of the farm operation (strata), historical data treatment procedure (Treatment), interviewer, and interactions (Strata*Trt) and (Trt.*Interviewer). The levels of each factor were the following. First, there were three stratum levels representing "small," "medium," and "large" acreage operations.[2]

The two treatment levels represented the control and direct use of previous grain stocks data. The number of interviewers in each State was 8 in California, 14 in Georgia, and 16 in Nebraska.

P-values are given, which represent the smallest level of significance that would have resulted in rejection of the null hypothesis. The smaller the P-value the more confidence one can place in that effect being important in explaining the variability in the data.

The results indicated the following. First, the overall model tests were, in general, highly significant (P-values <.01). This means that we can proceed to look at the individual components of the model, like the effects due to the historical data treatment procedure. Much of the variability in the models could be explained by the strata effect. This indicated that stratification was effective, which is analogous to block effects in the traditional design of experiment situation. Only in a few cases did the other model components contribute to the overall model variability. A discussion of these now follows.

---

[2] For example, in Nebraska, strata 50 (1-249 acres cropland), 54 (1-24,999 bushel capacity), 62 (hay acreage), 64 (livestock), and 69 (50-99 acres rye) comprised the "small" stratum. Strata 56 (250-499 acres cropland), 58 (25,000-49,999 bushel capacity), 60 (500-900 acres cropland), and 71 (100-999 acres rye) comprised the "medium" stratum. All other strata (66,68,72,80,82,96) which included 50,000 plus bushel capacity and 1000 plus acres cropland, comprised the "large" stratum. Random assignment of enumerators and treatments to the very many list frame strata would have resulted in small cell sizes and instability in the analysis of variance results.

Table 2. Analysis of variance results by crop storage item for Georgia 1/

Source of error	Degrees of freedom	Storage item: Capacity	Corn	Soybeans	All wheat
		P values			
Model	31	<.01*	<.01*	.16 2/	.27 2/
Strata	2	<.01*	<.01*	--	--
Treatment	1	.03*	.02*	--	--
Strata*Trt.	2	<.01*	.05*	--	--
Interviewer	13	.18	.43	--	--
Trt.*Interv.	13	.69	.08	--	--
Error	645				

1/ Weighted least squares (WLS) procedures were used, with weights proportional to the stratum variances.
2/ P-values for components of the model are not reported since the overall model was not significant (α=.10).
* Significant at the α=.10 level.

Table 3. Analysis of variance results by crop storage item for Nebraska 1/

Source of error	Degrees of freedom	Storage item: Capacity	Corn	Soybeans	All wheat
		P values			
Model	35	<.01*	<.01*	<.01*	.02*
Strata	2	<.01*	<.01*	<.01*	<.01*
Treatment	1	.90	.70	.74	.23
Strata*Trt.	2	.98	.86	.74	.40
Interviewer	15	.03*	.34	.52	.79
Trt.*Interv.	15	.25	.66	.22	.74
Error	1,020				

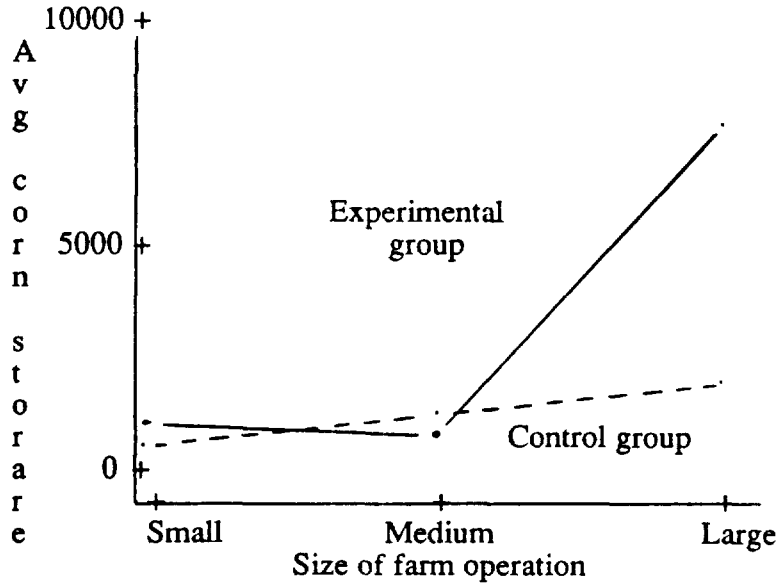
1/ Weighted least squares (WLS) procedures were used, with weights proportional to the stratum variances.
* Significant at the α=.10 level.

The primary interest in this study was to see if direct use of previous stocks data, on the average, caused responses to be biased when compared with the control group. In the absence of "true" stocks to precisely measure bias, we tested the following hypothesis which was the next best measure of "truth." That is, stocks should be significantly larger for the experimental than for the control group for corn, wheat, and soybeans. [3] No such direction could be tested for storage capacity, as they remain relatively constant in consecutive quarters.

Looking at treatment as the source of error in Tables 2 and 3, we see that only for storage capacity and corn stocks in Georgia were significant differences found in the weighted averages by historical data treatment procedure (α=.10). Also, the difference in the treatment means for stocks were in the direction hypothesized (382-bushel weighted average for the experimental group compared with 184 bushels for the control group). That is, respondents were influenced by their previously reported larger stocks, reporting higher levels than would have occurred had these prior data not been used. This is cause for concern.

Concentrating with these effects in Georgia, to just say that direct use of historical data had an effect when compared with the control would hide an important point: the difference was not constant across size of operation. That is, there was a treatment by strata interaction. Figure 1 graphs this interaction for average corn stocks.

Figure 1. Interaction of size of farm operation (strata) and historical data treatment procedure for average corn storage levels in Georgia



It appears that the expected effect of larger experimental group responses occurred

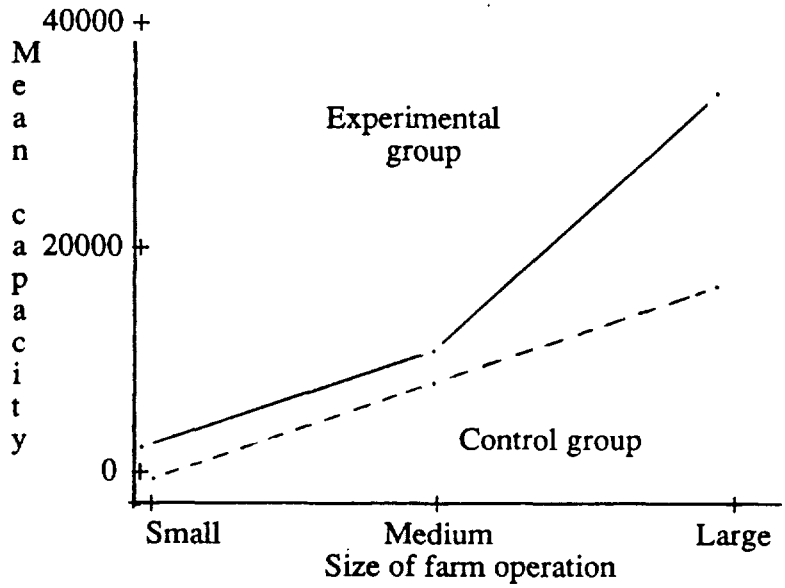
[3] Direct use of January stocks (stocks generally are larger in January, declining through April) in combination with respondents inclination to say "it is the same as last time," should produce, on the average, the higher levels for the experimental group.

just for the large operator stratum (p=.02), but not for the small (p=.21) and medium size (p=.64) operations. The p-values were calculated from statistical contrasts set-up to test for differences in treatments at each level of the strata effect (Steele and Torrie, 1980). They indicated that differences in the treatment procedure existed only in the large operator stratum (α=.10).

Reasons for this discrepancy other than the hypothesized treatment influence were next explored. First, the data were scanned for unusual or outlier observations. None were detected. Next, the frequency of "operator," "spouse, and "other" respondent reports in the two treatment groups was explored. This comparison revealed that the distributions were different, most notably with the percentage of "other" respondents in the large operator stratum. Approximately 13 percent of the experimental group responses came from the "other" respondents category, compared with 7.5 percent for the control group. The larger experimental group mean could be due to this unequal distribution, and not to any real differences, especially if farm manager operations, which are typically very large farms, composed most of the "other" respondent group. The exact cause of this interaction between historical data treatment procedure and size of the farm operation cannot be determined from this study.

A similar interaction relationship existed between the size of the farm (strata) and treatment procedure for grain storage capacity in Georgia (see Figure 2).

Figure 2. Interaction of size of farm operation (strata) and historical data treatment procedure for average storage capacity in Georgia



Here, the experimental group respondents from the large operator stratum reported much larger capacity compared with the control group respondents for both the medium (p=.05) and large farm operations (p<.01). No treatment difference could be detected in the small farm operation stratum (p=.71). The p-values again were derived from statistical contrasts. Differential proportions of respondents reporting among the treatment groups, again, may have contributed to these effects.

Analysis of direct use of previously reported stocks information in the current interview indicates the following. First, there was not overwhelming evidence that this

affected the stocks or storage estimates. However, just the presence of an effect in Georgia for corn stocks and capacity is enough to warrant concern. Second, the effect of using these previous data may be most evident with the large operations. Third, to be more conclusive, we need to control for respondent in these analyses so we can be sure differences are not due to certain respondent groups reporting more heavily than others.

Two other observations can be made from Tables 2 and 3. First, the results indicated very little mean differences by interviewer when we controlled for size of operation, treatment, and the other effects. This was expected given the highly controlled nature of the CATI environment, the fact that a research study was being done, and the type of questions asked were factual and closed-ended.

Second, the treatment by interviewer effect was not a consistent significant contributor to the overall model variability. That is, the same average response for the treatment procedure was obtained no matter which interviewer was considered. This is encouraging because it indicates the enumerators followed their interviewing procedures and instructions and were not contributing to any treatment differences that may have been present.

Comparison of Estimated Totals

Tables 4 and 5 present expanded State totals, coefficient of variation, and hypothesis tests for treatment differences in these estimated State totals for the various stocks items and capacity. Comparisons of this kind are beneficial because they measure the effect on estimated totals by change in survey procedures. P-values, given in the last column of Tables 4 and 5 indicate significant differences in the estimated totals by treatment procedure. A small P-value indicates the totals were different.

We note the following from Tables 4 and 5. For the respondent group that was given their January stocks report (experimental group), stocks were hypothesized to be biased upwards towards larger January levels. That is, we expected the experimental group totals to be larger than the control group totals. However, with grain storage capacity, these levels remain constant over the short run, and we would not expect much of this decline. Therefore, a two sided test was set up for testing for difference in the average storage levels.

Table 4. Estimated totals, coefficient of variation, and difference by historical data treatment procedure, and P-values for test of hypotheses--Georgia 1/

Crop	Control group Total	C.V.	Experimental group Total	C.V.	Difference	P-value
	1,000 bu.		1,000 bu.		1,000 bu.	
Corn	15,382	.20	27,270	.17	-11,889	.02*
All wheat	656	.39	1,327	.38	-670	.12
Soybeans	6,501	.26	2,978	.23	3,524	.98
Capacity	140,190	.10	129,319	.09	10,871	.56

1/ Hypothesis for crop storage tests was $H_o : \mu_{trt\ 1} = \mu_{trt\ 2}$ versus $H_a : \mu_{trt\ 1} < \mu_{trt\ 2}$. The alternative hypothesis for capacity was $H_a : \mu_{trt\ 1}$ not equal to $\mu_{trt\ 2}$.
* Significant as the α=.10 level.

Table 5. Estimated totals, coefficient of variation, and difference by historical data treatment procedure, and P-values for test of hypotheses--Nebraska. 1/

Crop	Control group Total	C.V.	Experimental group Total	C.V.	Difference	P-value
	1,000 bu.		1,000 bu.		1,000 bu.	
Corn	575,989	.08	610,444	.09	-35,455	.31
All wheat	32,815	.19	53,628	.23	-20,813	.07*
Soybeans	25,524	.13	25,590	.16	-66	.50
Capacity	1,286,556	.07	1,307,705	.07	-21,149	.86

1/ Hypothesis for crop storage tests was $H_o : \mu_{trt\ 1} = \mu_{trt\ 2}$ versus $H_a : \mu_{trt\ 1} < \mu_{trt\ 2}$. The alternative hypothesis for capacity was $H_a : \mu_{trt\ 1}$ not equal to $\mu_{trt\ 2}$.
* Significant as the α=.10 level.

The results in Tables 4 and 5 were not too unlike that of the earlier model testing. First, we see the same differences in the corn storage levels by treatment procedure for Georgia (α=.02). We do not note a difference in total storage capacity for Georgia as we did in the model testing section. In addition, there were statistical differences in the estimated totals for all wheat in Nebraska (α=.05) and Georgia (α=.12). Reasons for these differences are unknown.

Also of interest was the general trend of negative differences (experimental group totals larger) for the crop storage items. Combining this with the statistical differences noted above, we conclude there is additional evidence that respondents bias their answers to crop storage questions when given their previous report.

Comparison of Presence/Absence of Stocks by Treatment

In addition to interest in means and estimated totals, one would like to see if use of previous stocks information affected the reporting of presence or absence of crops in storage. That is, were there differences between treatment groups in the proportion of respondents answering "yes," or "no" to the question, "Is there any (crop) stored right now on the total acres you operate?" If differences existed one might suggest further research to determine if use of historical data improved recall, or forced a "yes" answer when no crops were in storage. This would be a validation study.

Table 6 presents P-values for chi-square tests comparing treatment procedure with presence/absence of storage. Small P-values indicated the proportions of respondents reporting crop storage varied by treatment procedure (see Appendix C for a complete discussion of independence testing in sample survey situations).

Table 6. Relationship of the presence/absence of crop storage and historical data treatment procedure by State and crop

Crop	California	Georgia	Nebraska
	Significance level		
Corn	.78	.87	.84
Soybeans	NA	.97	.78
All wheat	.06*	.06*	.97
Durum wheat	.97	NA	NA

NA - not applicable (item not estimated).
* - indicates significance at the α=.10 level.

We can see from Table 6 that the proportion of respondents reporting crops in storage was relatively unaffected by use of previous stocks data. Significance was found for all wheat in California and Georgia, but for no other crops across the three States. With all wheat in California, 7.8 percent of the control group respondents versus 2.2 percent of the experimental group respondents reported "yes" to the question of presence of this crop on the farm operation. The percentages in Georgia for all wheat were 0.9 percent for the control group and 2.5 percent for the experimental group respondents.

These results suggest that, at the bare minimum, the respondents report the existence of onfarm storage on their farm operation. Use of previous stocks information for that operation does not seem to effect this knowledge. The previous section addressed the issue of whether the levels were effected by prior survey information.

CATI Probing -- Storage and Stocks Changes Between Surveys

To assist in explaining reasons for change in stocks and capacity, a standard set of CATI probes was used (see Appendix A for examples of these questions). This analysis focuses on experimental group respondents who were asked to explain crop storage and storage capacity level differences of more than 100 bushels between surveys. [4]

Corn storage and storage capacity changes are discussed. This is because most of the January to April stocks activity (grain movement) is with corn. Also, accurate reporting of grain capacity has long been a concern within NASS.

It was previously noted that corn stocks generally decline during the January to April reference period as farmers sell, feed, and otherwise handle their crop. When reported stocks, instead, increase this may indicate a problem. The State Statistical Office will typically call respondents back prior to summarizing these data to inquire about unusual increases. This is done during the busy survey editing period and is cause for concern. Table 7 gives percentage of April responses for corn storage which were the same, below, and above the January levels. These data along with explanations for the stocks changes assist in interpreting these changes.

Table 7. Percent of April corn storage reports that were the same, above, or below the January corn storage reports (Experimental group respondents)

State	No change	April > Jan.	April < Jan.	Total	Experimental group response
	Percent				Number
California	91	3	6	100	233
Georgia	69	6	25	100	375
Nebraska	63	5	32	100	479
Total	71	5	24	100	1,087

From this table we see that in 71 percent of the cases no change occurred in corn storage levels, the levels decreased 24 percent of the time, and 5 percent of the responses ended with the April corn stocks larger than for January. The frequency distributions were similar for Nebraska and Georgia, but quite different for California.

Of concern, then, were the estimated 5 percent of respondents who said their corn stocks in April were larger than in January. CATI probing questions were used to uncover reasons for these increases. Table 8 presents these results.

[4] The 100 bushel difference was chosen arbitrarily small in order to obtain a response from every individual whose stocks or capacity may have changed. In practice, this difference would be larger so as not to burden the respondent with resolving minor changes.

Table 8. Frequency distribution of reasons for larger corn stocks levels in April compared with January

Reason for Increase	Freq.	Percentage of total
	Number	Percent
Bought corn	12	22
January response incorrect	19	36
"the corn is stored on a neighbors land, and in January they did not ask where it was stored"	1	
"he feels he estimated wrong in December"	1	
"not sure how much was stored January 6"	1	
"enumerator must have misunderstood... in December"	1	
"It is on his fathers farm he rents the bin from him; he did have it in December"	1	
"had the corn stored all winter; bushels unknown December 30"	1	
"was reported wrong"	2	
"did not include corn under loan"	1	
"this is not his corn so he said I just ignored it in December, but it is stored on his operation"	1	
"last survey was just where he lived"	1	
no comment	8	
Storing someone else's grain	1	2
Fed to livestock (mistakenly entered)	1	2
Other	20	38
"must not have understood respondent when he was called on January 3"	1	
"...estimated wrong in December"	2	
"grew his own & stored, no previous survey"	1	
"picked since last survey"	5	
"when the survey was done in Dec., he didn't know that he was to put in the sealed government stored corn.."	1	
no comment	10	
Total	53	100

As Table 8 indicates, a large majority of the "increases" may not have been real increases. Approximately 36 percent of the respondents who were given their January corn stocks and reported higher April levels said their January response was incorrect in some way. Also, some problems existed in getting accurate corn storage in responses coded as "other." Therefore, many responses could be termed problematic in the sense that onfarm stocks may not have changed. Further discussion of these problematic

responses now follows.

Table 9 shows the percentage of changes classified as problematic and valid from the total number of responses for which a change occurred for each stocks item and State. A problematic response was one where the accuracy of the January report was questioned by the respondent, or a comment was given indicating incorrect reporting in either time period. A "valid" response included the categories of buying and selling grains, grain fed to livestock, storing someone else's grain, and other comments which indicated correct reporting.

Table 9. Percentage of responses classified as problematic and valid out of the total number of responses for which a change occurred by State

Item/ State	Response		
	Problematic	Valid	Total
	Percent		
Storage capacity:			
California	90	10	100
Georgia	95	5	100
Nebraska	67	33	100
All States	80	20	100
Corn stocks:			
California	3	97	100
Georgia	17	83	100
Nebraska	13	87	100
All States	12	88	100
Soybean stocks:			
California	NA	NA	NA
Georgia	5	95	100
Nebraska	31	69	100
All States	21	79	100
All wheat stocks:			
California	MS	MS	MS
Georgia	6	94	100
Nebraska	32	68	100
All States	MS	MS	MS

NA - Not available (crop not estimated).
MS - Data were missing.

Of interest in this table was the large discrepancy in percentage of problematic responses between stocks and storage capacity. More than two-thirds of the responses (67-95 percent) could be classified as problematic when capacity changed from January to April; the change may not have been real change. On the other hand, less than one-third of the responses (3-32 percent) could be classified in this way for crop storage items. These results suggest that quarterly changes in storage capacity are not

being scrutinized adequately by interviewers and field office personnel. More emphasis needs to be given in this area. Storage capacity is generally stable over the short run, while stocks are built up and regularly taken off the farm.

CATI has additional advantages in allowing enumerators to optionally key comments in the instrument at any point, or require comments in specific places. Comments are now summarized that explain what some of the problematic responses were. Most problematic responses given were that "the historical report was incorrect" with no other information. A summary of other responses which indicated there were problems is given next. First, the biggest problem seemed to be identifying what stocks or capacity to report. For capacity, problems included respondents not reporting capacity on rented land or land operated for a relative, respondents not reporting bins that were usable but not in use at the time, and a general "misunderstanding" on what to include such as temporary capacity. For stocks, problems included not reporting grain in rented bins on other farms, grain kept for feed, grain stored under government programs, and grain belonging to someone else but stored on the respondent's operation.

The comments also included several remarks that the respondent just forgot to report some stocks or capacity, several comments indicating that the respondents were unsure of their grain capacity or stocks in January, and one comment stating that "he doesn't know where we got the figure in January, but he had the wheat then." Finally, two respondents said the enumerator incorrectly recorded the information in January. The January Survey was not done with CASES.

Change in Respondents and Its Effects

Was the change in respondent between surveys responsible for the increase or decrease in stocks or storage capacity, or did the levels just change? Analysis was restricted to the same group defined in the previous section. This was TREATMENT 2 respondents whose corn stocks or storage capacity changed by more than 100 bushels between the January and April surveys.

The responses of problematic and "valid" presented in the previous section were compared against respondent categories "same-same," and "other." The "same-same" respondent category was where the respondent was the same in both surveys (operator-operator, spouse-spouse, other-same other). In the majority of cases, the operator reported in both surveys. This is because of NASS survey procedures which instruct the interviewer to contact the operator where possible. The "other" category covered all other combinations of reporting (operator-spouse, spouse-operator, spouse-other, other-spouse, other-operator, operator-other, other-different other). More meaningful comparisons involving the detailed groups of operator-operator, spouse-other, etc., were not possible because cell counts were too small (often zero or fewer than five).

Table 10 reports the results of tests of independence in the two classifications (see Appendix C for a complete discussion of independence testing in sample survey situations). We expected that more problematic responses would occur if someone different responded ("other") between surveys. The # sign in Table 10 indicates significance at the α=.10 level, and the existence of this relationship.

This table shows that the change in respondent in many cases played an important role in changes in stocks and capacity. What was generally found was that problematic

responses were more abundant than "valid" responses in the respondent group that remained unchanged between surveys. This occurred for corn stocks and storage capacity in California and storage capacity in Nebraska (α=.10). For example, the percentage distribution for corn stocks in California was 65 percent problematic when the respondent changed, compared with 0 percent problematic when the respondent remained unchanged. For storage capacity in Nebraska, 96 percent and 62 percent were problematic for the "other" and "same-same" respondent groupings.

Table 10. Significance levels for the comparison of the proportions of valid grain stocks responses by "same-same" and "other" respondent categories

	P-values		
Crop	California	Georgia	Nebraska
Corn	.07#	.22	.15
Capacity	<.01#	.29	<.01#

Significance at the α=.10 level, and larger percentage of responses classified as problematic for the "other" respondent group compared with the "same" group.

Enumerator Post-Survey Comments

After the close of calling, each enumerator was asked to complete a post-survey evaluation (see Appendix D). All 8 California and all 16 Nebraska enumerators responded, while only 7 out of approximately 15 Georgia enumerators completed this evaluation. The Georgia enumerators who failed to respond were not refusals, but just failed to receive a post-survey questionnaire. There is no reason to believe their responses were different than those responding. A summary of enumerator responses and comments now follows. Note the use of the terms VERSION 1 and VERSION 2. These correspond to the control and experimental groups.

The enumerators were first asked to evaluate farmer reaction to direct reference to the January stocks report (experimental group) and note any objections. Across all States, 81 percent of all enumerators stated that, "farmers you interviewed using VERSION 2, in general, never objected to our use of the January ISP stocks data for the April survey." The remaining 19 percent stated "very rarely" were there any objections. Comments written by the enumerators were that only a couple of inquires were made, but no actual objections. One farmer was impressed that we knew he reported in January, another was concerned with where the data came from until informed it was reported by him previously, and another farmer claimed no one contacted him previously.

When the enumerators were asked to choose between versions, 3 enumerators preferred VERSION 1 (no use of January stocks data), 15 preferred VERSION 2 (direct use of January stocks data), and 13 had no preference. Five out of the 7 Georgia enumerators reported no preference, while only 6 of 16 Nebraska, and 2 of 8 California enumerators reported no preference. Of the 10 Nebraska enumerators reporting a preference, nine,

or almost all, preferred using the historical data. Four California enumerators out of the six with a preference, chose direct use of historical data. These counts may indicate an effect due to office CATI experience. The Georgia enumerators had used CATI only once or twice before this survey, and may have been overwhelmed with the process of learning CATI. The Nebraska and California enumerators, however, had extensively used CATI.

Twenty-three of the 30 interviewers felt some improvement could be made in the wording of VERSION 2 if used again. Comments on the wording of the questions were mixed. Some suggestions were for minor wording changes. Others suggested reorganizing the stocks and capacity questions.

The final two questions dealt with whether the date of the previous interview and knowledge of the previous respondent were useful. Recall that these two pieces of information were included in CATI screens to assist the enumerator in answering questions about where, how, and when the previous stocks data were obtained (see Appendix A). We do not, however, present the frequency distributions for the question dealing with the use of the previous respondent. It was evident that this question was misinterpreted. However, there were some comments of value for this question. One enumerator said that use of the previous respondent put the interviewer and respondent on a one-to-one basis. Other enumerators said "it was an absolute must to have the information available," "it was helpful because sometimes I was asked who gave out the information," and "the farmers feel we are compiling information on him personally."

A mixed response was obtained for the question on use of the date of the previous interview. Ten enumerators considered having the date of the previous interview to be very useful, eight somewhat useful, seven had a neutral opinion, and six considered it not useful. Three of eight California enumerators, and four of eight Georgia enumerators considered the date of the previous interview very useful or somewhat useful. This compares with 11 of 16 in Nebraska. An explanation for this may be the following. Little or no stocks or storage typically existed for the sample units in Georgia and California, and enumerators had little use for the date of the previous interview if it was always zero. This was not the case in Nebraska, as positive crop storage and capacity was common.

Enumerators who preferred use of the date of the previous interview commented that "it assisted the grower in recalling the information," "having the data available broke the ice," "he knew when he bought and sold," "it is essential to have when historical data is referenced in that way," "can give exact date in case they have forgotten," and "it was easier for the farmer to think back to the date given as to how much grain was left."

Those enumerators who were neutral or objected to the use of the date of the previous interview commented that "it simply did no good," "the actual date made the question cumbersome," "it did not improve the quality of the data or quicken the speed of the interview," "no one said anything about the date," and "most respondents did not remember anyway."

DISCUSSION

The research study set out to measure the effect of directly providing previously collected grain stocks and storage capacity data to every respondent in the current CATI grain stocks portion of the ISP interview. The exact effect could not be determined since "true" grain storage and capacity levels were not available. However, because of good prior knowledge of the relationship in the change in grain storage levels between January and April, we can be fairly certain of the conclusions that were made.

It was hypothesized that the average grain stocks response for the experimental group would be higher than for the control group. In addition, the experimental group's average grain storage capacity should be either above or below that of the control group. The analyses provided evidence for support of these hypotheses. The evidence comes from model testing, where these relationships were found for corn storage and storage capacity in Georgia, and from the analysis of the estimated totals for corn storage in Georgia, and all wheat in Georgia and Nebraska.

The fact that enumerators had little flexibility in the way historical data were referenced and knew this was a research study, where strict attention to survey procedures was a requirement, may have hidden even more evidence of these biases. On the other hand, fewer differences may have existed had there been control over the choice of respondent (operator, spouse, other). It was shown that the experimental group had a larger percentage of responses from the "other" respondent group in the large operator stratum than occurred for the control group.

This study suggests that NASS not use previously reported grain stocks information for every respondent in the quarterly grain stocks surveys using CATI. No recommendation can or should be made for other types of data collection activity. However, as a general rule, it would not be advisable, in Quarterly Agricultural Surveys, to allow enumerator access to prior grain stocks responses in less supervised modes of data collection such as with non-CATI and personal interviews. The cited studies have shown that responses can be easily influenced even with specific interviewer instructions on how to use prior information.

These results should not discourage believers in use of previous survey data. There are many applications where its use is important, and we should now begin to study these. Pafford (1986), for example, previously suggested use of historical data as an online editing tool in CATI. The application would also apply to quarterly grain stocks surveys. The procedure would be to have a standard CATI check of the current response with the previous stocks response, and a standard probe come forward when it deviates "too much," or is outside some prespecified range. These prespecified ranges should be large enough to allow for unbiased responses. Further uses of prior survey data could be for a "base" in a balance sheet data collection procedure. Balance sheet procedures start with some base number, and ask respondents for deletions and additions since the base. Any procedures like these should be researched, just like the "direct use" method was studied in this paper. Recommendations for research along these lines are outlined in Appendix E.

Other objectives of this paper were to evaluate effects due to the respondent, size of operation, and interviewer. In terms of respondent effects, it was discovered that interviewing different respondents between quarters created many more problems in identifying "true" level changes. This specifically suggests that even more emphasis should be given in contacting the same person, preferably the farm operator, in grain

stocks surveys. The application to CATI is ideal. The enumerator could be reminded up-front in the CATI interview (for example, at the time the respondent is selected) who the prior respondent was and be prompted to obtain a response from the operator. In addition, information on the size of the farm operation, solely in terms of an indicator, would be useful. Interviewers would be instructed, with the largest operators, to be very careful in selecting a respondent, even if it means making a callback. These changes should improve the quality of NASS's survey data, particularly with grain stocks.

In terms of size of the farm operation, it was discovered that large farms tended to be influenced more by direct use of prior stocks data than the operations of smaller size. These biases can be controlled by avoiding direct use of prior survey data as used in this study. This is not to say that prior survey data not be used at all for large farm operations. At times it is very important, especially when reporting unit problems exist. The question, "Which farms, tracts, parcels, businesses, subsidiaries do you want?" can be more answerable with prior survey data. These specific uses should be studied.

Finally, little evidence of interviewer effects were observed in this April grain stocks survey. This finding was consistent with other reports for telephone interviews, and was expected given the structured nature of the CATI interview and the experimental setting for this research project. This particular result is encouraging because it suggests this survey was not influenced differentially by the different interviewers. Unequal workloads then would not be as much of a concern if one knows that the interviewers "perform" about the same on the average.

RECOMMENDATIONS

The following is a list of recommendations:

(1) Direct reference to the prior quarter's grain stocks response for every respondent in the current grain stocks CATI interview should be avoided. This study has shown that, on the average, farmers will give biased responses if the data are used in this manner.

(2) More emphasis needs to be given to maintaining consistency in the reporting unit between operations sampled in consecutive surveys. Prior respondent should be brought forward for use during the CATI interview for overlapping sample units. Enumerators should use these data and attempt to contact the operator where possible. Immediate implementation is needed.

(3) Other ideas on use of prior grain stocks data in CATI should be considered. A preliminary project proposal is given in Appendix E that recommends a March 1988 study in all 15 CATI States to compare the current operational procedure, direct use of previous stocks when there are "large discrepancies," and a balance sheet questionnaire approach.

REFERENCES

Bailar, B.A. (1968). "Recent Research in Reinterview Procedures," JASA, Vol. 63, 41-63.

Bosecker, R.R. (1977). "Data Imputation Study on Oklahoma DES," Staff Report, National Agricultural Statistics Service, U.S. Department of Agriculture, October.

Draper, N.R., and H. Smith (1966). Applied Regression Analysis, John Wiley & Sons, Inc., New York.

Fuller, W.A. (1973). "Regression Analysis for Sample Surveys," Presented at the International Statistical Institute, Vienna.

Fuller, W.A. (1975). "Regression Analysis for Sample Surveys," Sankhya, Series C, Vol 37, 117-132.

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

Groves, R.M., L.J. Magilavy, and N.A. Mathiowetz (1981). "The Process of Interviewer Variability: Evidence From Telephone Surveys," Proceedings of the American Statistical Association, Section on Survey Research Methods, 438-443.

Hansen, Morris H., William N. Hurwitz, E.S. Marks, and W.P. Maudlin (1951). "Response Errors in Surveys," JASA, Vol. 46, 147-90.

Hansen, M.H., W.N. Hurwitz, and L. Pritzker (1964). "The Estimation and Interpretation of Gross Differences and the Simple Response Variance," in C.R. Rao, ed. Contributions to Statistics Presented to Professor P.C. Mahalanobis on the Occasion of his 70th Birthday, Calcutta: Pergamon Press, Ltd., 111-136.

Hansen, M.H., and E.S. Marks (1958). "Influence of the Interviewer on the Accuracy of Survey Results," JASA, Vol 53, 635-655.

Hidiroglou, M.A. (1974). "Estimation of Regression Parameters for Finite Populations," Unpublished Ph.D. thesis, Iowa State University, Ames, Iowa.

Kish, Leslie (1962). "Studies of Interviewer Variance for Attitudinal Variables," JASA, Vol. 57, 92-155.

Mahalanobis, P.C. (1946). "Recent Experiments in Statistical Sampling in the Indian Statistical Institute," JASA, Vol. 109, 325-70.

Nealon, Jack, and David Dillard (1984). "Response Comparisons Between Husbands and Wives for Farm Characteristics," SRS Staff Report Number AGES830617, National Agricultural Statistics Service, U.S. Department of Agriculture.

O'Muircheartaigh, Colm A. (1986). "Correlates of Reinterview Response Inconsistency in the Current Population Survey," Proceedings of the Second Annual Research Conference, U.S. Bureau of the Census, March 23-26, Reston, Va. 1985.

Pafford, Brad (1986). "Response Errors in NASS Surveys: The Effect of Using Previous Survey Data in the 1985 California Fall Acreage and Production Survey," Staff Report Number SF&SRB 86-99, National Agricultural Statistics Service, U.S. Department of Agriculture.

Pafford, Bradley V. (1987). "Changing Survey Reference Dates: An Analysis of the June 1986 Bridge Survey," Staff Report Number SRB-87-07.

PCCARP (1987) Computer software written under the direction of W.A. Fuller, Statistical Laboratory, Iowa State University, Ames, Iowa.

Steele, Robert G.D. and James H. Torrie (1980). Principles and Procedures of Statistics: A Biometrical Approach, 2nd Edition, McGraw-Hill, New York.

Steiner, Mike (1980). "Study of Hog Balance Sheet and Alternative Ways to Ask Farrowing Questions," Working Paper, National Agricultural Statistics Service, U.S. Department of Agriculture.

Tortora, Robert D. (1980). "The Effect of a Disproportionate, Stratified Design on Principal Component Analysis Used for Variable Elimination," Staff Report, National Agricultural Statistics Service, U.S. Department of Agriculture.

Warde, William D. (1986) "Examination of the Effect of the Respondent and Collection Method on Survey Results," Staff Report Number SF&SRB 86-94, National Agricultural Statistics Service, U.S. Department of Agriculture.

Warde, William D. (1987) "Contacting the Farm Operator by Telephone: A Re-examination," NASS working paper available from the author.

APPENDIX A: CATI INSTRUMENT QUESTIONS

This appendix presents the general form of the questions for each historical data treatment group. There was the control group, where no historical data were referenced, and the experimental group where direct reference was made of these data. The questions are only given for corn stocks and storage for each treatment group. Many of the automatic edit checks and their associated CATI questions, such as when the sum of the grain stored exceeds capacity, are not presented.

Control group question format - corn stocks

Please account for the whole grains (and oilseeds) stored on
the total acres you operate, whether for feed, seed or sale.
They can belong to you or someone else -- or be stored under a
government program (loan, farmer owned reserve or CCC).

Is there any WHOLE GRAIN corn stored right now on the total
acres you operate?

<1> YES
<2> NO
<3> DON'T KNOW

====>

Control group question format - capacity

What is the TOTAL STORAGE CAPACITY of all bins,
cribs, sheds, and other structures normally used to store
whole grains on the total acres you operate?

<0 - 5000000> tons
<n> NO ANSWER

====>

Experimental group question format - corn stocks

Please account for the whole grains (and oilseeds) stored on
the total acres you operate, whether for feed, seed or sale.
They can belong to you or someone else -- or be stored under a
government program (loan, farmer owned reserve or CCC).

********** PREVIOUS RESPONDENT: [fill with name] **********

Our records show that on [fill with previous date] your operation
reported [fill with previous stocks] tons of CORN being stored.

Is there any corn stored on your operation?

<1> YES
<2> NO
<3> DON'T KNOW

====>

Experimental Group question format - reason for change in corn stocks

If the current stocks varied by more than 100 bushels from the January response the following question was asked:

Why is the current stocks of CORN [above/below] the
[fill with previous date] level?

<1> [fill with previous date] report incorrect -- ENUM: Ask
for correct [fill with previous date] stocks and enter
after "specify."
<2> Fed to livestock
<3> Storing someone else's grain
<4> other

====>

Experimental group question format - capacity

Our records show that on [fill previous date] you operation reported to us a GRAIN STORAGE CAPACITY of [fill with previous capacity] tons.

What is NOW the total CAPACITY of all bins, cribs, sheds, and other structures normally used to store whole grains on the total acres operated by you?

********** PREVIOUS RESPONDENT: [fill with name] **********

<s> SAME as previous quarter
<0 - 5000000> tons
<n> NO ANSWER

====>

Similar to the stocks questions, if the current capacity varied by more than 100 bushels from the January response the following question was asked:

Experimental group question format - reason for change in capacity

What is the reason for the change in capacity from [fill with previous date]?

[fill with previous date] CAPACITY == [fill with prev. capacity] tons
CURRENT CAPACITY == [fill with current capacity] tons

**** to change current capacity -- backup :b and change :ca)

<1> [fill with previous date] report incorrect -- ENUM: Ask for correct [fill with previous date] stocks and enter after "specify."
<2> Added facilities
<3> Destroyed facilities
<4> Sold or rented facilities
<5> Other[specify]; please explain

====>

APPENDIX B: WEIGHTING AND MODEL EXPECTED MEAN SQUARES

This appendix presents the precise form of the model used to analyze effects due to the size of operation (strata), historical data treatment procedure, interviewer, and their interactions. The interaction effects left out of the model were strata by interviewer, and strata by treatment by interviewer. These were not considered important sources of variation. In addition, the cell counts for the three factor interaction were too small (much less than 30) to provide stable test results. All effects were considered fixed. That is, the factor levels were chosen by the researcher and generalization was just to these sets of treatments, operations, and enumerators. Also shown is the weighting scheme used in this model. Finally, the expected mean squares are presented to show how the tests of significance were constructed.

Model

$$Y_{ijkl} = \mu + S_i + T_j + ST_{ij} + I_k + TI_{jk} + \varepsilon_{l(ijk)}$$

$$i=1,2,3 \quad j=1,2 \quad k=1,2,...,w$$

where,

Y_{ijkl} = grain stocks report for the l-th unit in the i-th stratum, j-th historical data treatment, and k-th interviewer,

μ = overall mean effect,

S_i = the effect of i-th stratum (fixed effect),

T_j = the effect due to the j-th historical data treatment group (fixed effect),

ST_{ij} = the interaction effect of using the j-th treatment in the i-th stratum

I_k = the effect of the k-th interviewer (fixed effect),

TI_{jk} = the interaction effect of using the j-th treatment for the k-th interviewer,

$\varepsilon_{l(ijk)}$ = the random effect within the i-th, j-th, k-th cell.

Weighting

Weighted least squares (WLS) were used with weights proportional to the strata variances. One of the assumptions of analysis of variance is that the error terms are distributed normally, with mean zero and variance $\sigma^2 I$. That is, each observation has equal variance. If this is not so, the data must be transformed. In this study, observations in each strata have equal variance, but not across strata. Therefore, each observation was weighted inversely proportional to their stratum variance. We present this more formally.

From Draper and Smith (1966) we know that for the model $\mathbf{Y} = \mathbf{XB} + \mathbf{e}$, if $\mathbf{e}$ is distributed as $N(0, V\sigma^2)$ the solution to the normal equations for $\mathbf{b}$ is $\mathbf{b}=(X'V^{-1}X)^{-1}X'V^{-1}Y$. We let $\mathbf{V}$ (n x n matrix) equal a diagonal matrix with diagonal elements equal the proportion of the total sample variance (σ^2) accounted for by that stratum. Then $V\sigma^2$ and V equaled the following:

$$V\sigma^2 = \begin{vmatrix} \sigma_1^2 & & & & & \\ & \ddots & & & 0 & \\ & & \sigma_1^2 & & & \\ & & & \sigma_2^2 & & \\ & & & & \ddots & \\ & 0 & & & & \sigma_3^2 \end{vmatrix} \qquad \mathbf{V} = \begin{vmatrix} \sigma_1^2/\sigma^2 & & & & & \\ & \ddots & & & 0 & \\ & & \sigma_1^2/\sigma^2 & & & \\ & & & \sigma_2^2/\sigma^2 & & \\ & & & & \ddots & \\ & 0 & & & & \sigma_3^2/\sigma^2 \end{vmatrix}$$

The subscripts refer to the stratum. $V\sigma^2$ is a n X n matrix, with n sample observations. The off diagonal elements were zero since the strata were independent.

Expected Mean Squares

Typical expected mean squares for the model tests were as follows. Note that the correct denominator for the F-tests was the error term in all cases since all effects were considered fixed in the model.

Source	Expected Mean Squares
Strata	$\sigma_\varepsilon^2 + \phi$(Strata)
Treatment	$\sigma_\varepsilon^2 + \phi$(Trt,Strata*Trt,Trt*Interv)
Strata * Treatment	$\sigma_\varepsilon^2 + \phi$(Strata*Trt)
Interviewer	$\sigma_\varepsilon^2 + \phi$(Interv,Trt*Interv)
Treatment * Interviewer	$\sigma_\varepsilon^2 + \phi$(Trt*Interv)
Error	σ_ε^2

APPENDIX C: 2X2 TESTS OF INDEPENDENCE

The tests of independence for the contingency tables in the sections titled **Comparison of Presence/Absence of Stocks by Treatment**, and **Change in Respondents and Its Effects** were made taking into consideration the survey design effects. Computer software from Iowa State University Statistics Laboratory, called PCCARP (1987), was used. This software implements statistical calculations outlined in Fuller (1973, 1975), and Hidiroglou (1974). A brief discussion of the tests will now follow as applied to analyzing the effects of change in respondent.

We let,

r = row classification variable (response), r = 1,2,3.

1 = problematic 2 = valid
3 = not in the domain (Experimental group respondent who's stocks or capacity did not change by more than 100 bushels)

c = column classification variable (respondent) c = 1,2,3.

1 = "same-same" 2 = "other"
3 = not in the domain (Experimental group respondent who's stocks or capacity did not change by more than 100 bushels)

w_{ij} = sampling weight (expansion factor) for the j-th element in stratum i.

$$Y_{(rc)ij} = \begin{cases} 1 \text{ if the j-th element in the i-th stratum falls into the rc-th cell} \\ 0 \text{ otherwise} \end{cases}$$

Then, the estimated total, and proportion for the rc-th cell are:

$$Y_{(rc)} = \sum_{i=1}^{3} \sum_{j=1}^{n_i} w_{ij} Y_{(rc)ij}, \text{ and } P_{rc} = Y_{..}^{-1} Y_{rc}$$

Under simple random sampling assumptions and under the null hypothesis (of independence), the ordinary Pearson chi-square statistic (below) is distributed as a chi-square random variable with (r-1)(c-1) degrees of freedom

$$X_P^2 = n \sum_{r=1}^{R} \sum_{c=1}^{C} P_{r.}^{-1} P_{.c}^{-1} (P_{rc} - P_{r.}^{-1} P_{.c}^{-1})^2$$

However, for samples other than simple random sampling (disproportionate stratified simple random sampling in this study), the degrees of freedom must be estimated, and an approximate statistic computed for tests of independence.

If the design effect, which is the ratio of variance of under the sample design to the

simple random sampling variance, is 1 the Pearson chi-square statistic would be appropriate.

From PCCARP (1987), the approximate test statistic equals F_d which is a function of X_P2. F_d is F-distributed with an approximated numerator degrees of freedom h_f and denominator degrees of freedom, d, equal to

$$d = \sum_{i=1}^{3} (n_i - 1),$$

where n_i equals the number of elements in stratum i. The design effect is estimated by the quantity,

$$(X_P^2)^{-1}(R-1)(C-1)F_d .$$

APPENDIX D: ENUMERATOR POST-SURVEY QUESTIONNAIRE

NAME:_____

INTERVIEWER EVALUATION OF 1986 APRIL ISP SURVEY

Nebraska - March 1986

Instructions: Please help us evaluate this survey by completing the following questions as honestly and as thoughtfully as possible. Your response will be kept confidential!

Filling out the questionnaire will be very helpful in determining good ways to use historical data for future surveys.

For each item, check the appropriate response and give any comments you feel are important to the question.

(1) How much CATI experience have you had?

a. No previous CATI experience
b. At least one previous CATI survey worked

(2) How much regular (non-CATI) telephoning experience do you have?

a. Less than 6 months
b. 6 months - 1 year
c. 1 - 2 years
d. More than 2 years

(3) Do you have a farm background? (Raised or worked on a farm or gained a fairly good knowledge of farm activities from others who are farmers)

a. Yes
b. No

(4) How difficult would you say it was for you to learn to interview using CATI?

a. Very Difficult
b. Somewhat Difficult
c. Somewhat Easy
d. Very Easy

****
****
* *
* TO GET ON COMMON FOOTING LET US MENTION AGAIN THE TWO WAYS *
* PREVIOUS JANUARY ISP DATA WERE USED. THE FOLLOWING QUESTIONS*
* WILL REFER TO THESE VERSIONS GIVEN BELOW. *
* *
* VERSION 1. You did not see any previous figure. Case IDs *
* began with the number 1. (For example, 12101). *
* *
* VERSION 2. The previous data was worded into the question.*
* The questions read like "Our records show that *
* on (date) your operation reported (bushels) of *
* CORN being stored? Is there any CORN now being* *
* stored on this operation?" *
****
****

(5) Circle the appropriate response that best completes this sentence?

Farmers you interviewed using VERSION 2, in general, _____ to our use of the January ISP stocks data for the April survey!

a. Strongly objected
b. Somewhat objected
c. Very rarely objected
d. Never objected

COMMENTS:

(6) With respect to the preceding question, estimate the number of times the respondents openly objected for ALL the calls you made.

a. 1
b. 2-5
c. 6 - 10
d. 10+

COMMENTS:

(7) If you had to choose between VERSIONS 1 and 2 which would you prefer?

a. VERSION 1
b. VERSION 2
e. No preference

(8) What is it you liked about this VERSION (the one selected in the last question) above all the rest?

COMMENTS:

(9) Could the wording on the VERSION 2 questions be improved?

a. Yes I would change the wording (give an example of your revision).
b. No I would not change the wording

COMMENTS:

(10) Was it useful to have information on the January ISP RESPONDENT available for the interviews that you did?

a. Very Useful
b. Somewhat Useful
c. Neutral
d. Somewhat Not Useful
e. Not Useful

COMMENTS: (What did you like or dislike about this?)

(11) Was it useful to have the DATE of the January ISP interview available for the calls that you did?

a. Very Useful
b. Somewhat Useful
c. Neutral
d. Somewhat Not Useful
e. Not Useful

COMMENTS: (What did you like or dislike about this?)

(12) ARE THERE ANY OTHER COMMENTS YOU FEEL WOULD BE OF INTEREST FOR EVALUATING THIS SURVEY? (If so please explain!)

APPENDIX E: PRELIMINARY RESEARCH PROPOSAL

PURPOSE

The purpose of this study is to evaluate alternative methods of using prior survey data in the CATI telephone interview environment. The results outlined in this paper and that of Pafford (1986) suggest that methods other than direct use of previously collected survey data be considered.

I propose to study two methods. First, prior data would be used in an online editing fashion. Current responses would be instantaneously checked against any prior response during the initial CATI interview, with standard probing questions used to resolve discrepancies. The second method involves a balance sheet format. Prior data would act as a base, and changes since the base would be asked.

JUSTIFICATION

The justification for use of prior data is well documented in this paper and that of Pafford (1986). Two methods have been evaluated by the author so far. These include (1) direct use (this research paper and Pafford, 1986), and (2) use of prior data in the margins of the CATI screen (Pafford, 1986). The results suggest that these methods bias the respondents answers. Therefore, alternative procedures need evaluating.

Use of prior stocks data as an editing tool has been suggested by many individuals. Applications to crops and livestock exist as well. The Nebraska State Statistical Office recently called the author, suggesting previous quarter stocks data be made available for their enumerators to use in some quarterly stocks surveys. They found, for the March 1987 Quarterly Agricultural Survey (QAS), that many operations had to be called back a day or two later after computer edits revealed stocks increased from the January 1987 report. When called back, the respondent invariably corrected the initial response. The state office personnel suggested that, for the quarters where the largest decline in stocks is expected, prior stocks data be brought forward in the CATI interview when stocks increase. This would save recalling many operations.

Discussions of balance sheet data collection methods are not new. These have been suggested in an attempt to reduce the bias of the operational procedure estimates. Pafford (1987), for example, in his study comparing two reference date data collection procedures, noted that a large proportion of interviewers found the hog and pig questions very hard for farmers to complete accurately.

Steiner (1980) studied a hog and pig inventory balance sheet approach in Nebraska. He concluded that this method was less reliable than the operational procedure. Respondents had difficulty in reporting gains and losses (noted from enumerator comments and farmer response rates), more difficulty over the phone versus by personal interview, and estimates were generally larger. Who the respondent was between quarters seemed to be important, as well as the time of completing the previous report to the first of the month. Only 50 samples were used for this analysis, no statistical tests were made, and the respondent may have been biased towards the operational procedure since it preceded the balance sheet approach in all cases.

Application of balance sheets to grain stocks has not been formally studied to the author's knowledge. It is expected that the amount of detail required is much less than

for hog and pig inventory, and accuracy should be much better.

STUDY

The March 1988 QAS should be chosen and a study initiated in as many CATI States as needed to detect a prespecified difference in the three procedures outlined above. Split sample testing techniques would be used to evaluate the balance sheet, online editing, and operational approaches for selected items. The balance sheet approach should be restricted to grain stocks.

Techniques similar to Steiner's (1980) of asking the operational and balance sheet approaches during the same interview could be applied and comments solicited from the respondent. However, more controls such as randomly assigning the order of the approaches need to be applied.

Further details on the study design, including cost estimates, can be formulated if there is general interest in this topic area.