United States Department of Agriculture

National Agricultural Statistics Service

Research and Development Division
Washington DC 20250

RDD Research Report
Number RDD-11-05

September, 2011

National Agricultural Statistics Service: Research and Perspective on the Use of Previously Reported Data During Data Collection

Michael W. Gerling
HoaiNam N. Tran
Terry P. O'Connor

National Agricultural Statistics Service: Research and Perspective on the Use of Previously Reported Data During Data Collection

Michael W. Gerling, HoaiNam N. Tran, Terry P. O'Connor[1]

Abstract

The National Agricultural Statistics Service (NASS) surveys farmers and ranchers across the United States and Puerto Rico in order to estimate crop production and number of livestock, to assess production practices, and to identify economic trends. This report is a culmination of NASS research over the years on the use of previously reported data (PRD) in an attempt to improve the data collected.

Key Words: Agriculture, Data Collection, Previously Reported Data, Surveys

1. INTRODUCTION

Effectively and efficiently utilizing previously reported data (PRD) to improve data quality and reduce respondent burden continues to generate much discussion within the National Agricultural Statistics Service (NASS). The potential benefits of expanding the use of PRD are to improve data consistency and reduce respondent burden. However, if PRD is used improperly, it may increase respondent burden and measurement error, leading to biased survey results. This document provides an overview of NASS past research and current operational uses of PRD during data collection. This topic was previously discussed at the March 2006 NASS Program Planning Council (see Appendix A).

2. A COMPILATION OF NASS' PREVIOUSLY REPORTED DATA RESEARCH

The National Agricultural Statistics Service has conducted several studies over the years focusing on previously reported data. Table 1 summarizes the history of this research. Gottschall (2009) includes a summary of NASS uses of PRD and recommendations addressing the potential future directions of PRD within NASS.

[1/] Michael W. Gerling and HoaiNam N. Tran, are Mathematical Statisticians and Terry P. O'Connor is Deputy Chief of the Survey Methodology and Technology Branch for the National Agricultural Statistics Service - Research & Development Division, located at 3251 Old Lee Highway, Fairfax, VA 22030.

Table 1: Previously Reported Data Research within the National Agricultural Statistics Service

Unpublished Previously Reported Data Research - 2010, Tran, H., Gerling, M., Research and Development Division, United States Department of Agriculture, National Agricultural Statistics Service.

Based on the report Using Previously Reported Acreage in Data Collection, the authors decided to see if additional items (grain storage capacity, land owned, land rented to, and land rented from) were examined to determine the stability of the individual operations' values between June and September Agricultural Surveys. The authors decided to discontinue the research when they found that these values, which have been thought to be fairly stable, turned out to be very unstable even over the course of a few months. Percentage changes between quarters for storage capacity percentage changes range from 0.2% to 38.2%. Between-quarter percentage changes for land owned, land rented from, and land rented were extreme, ranging from 0.2% to 3917%.

Using Previously Reported Cropland Acreage in Data Collection, - 2009, Tran, H., Gerling, M., Research and Development Division, RDD-09-01, United States Department of Agriculture, National Agricultural Statistics Service

In an effort to reduce respondent burden, NASS' Research and Development Division examined the relationship between the June and September Agricultural Surveys' data to determine whether certain questions (cropland, land owned, land rented to, land rented from, and storage capacity) could be dropped from the September survey if the respondent had already answered the same question in June.

Five states, representing high, middle, and low agricultural production were selected for a detailed analysis of the potential results of carrying forward the previously reported data. Cropland was the first item analyzed since it is asked on most surveys that NASS conducts and is one of the questions that respondents complain about being asked repeatedly. The June and September Agricultural Survey questions on cropland acreage operated were examined and the differences between June and September responses for the same operation were measured.

The initial results for cropland data varied by state, so to obtain a broader view across states, a less detailed analysis based solely on the change in aggregate results was done for all states. Outliers were reviewed and found that most of the extreme outliers were either data collection or data editing errors. Excluding these erroneous records from the analysis, it shows that June's total cropland could be carried forward to the September for four of the five states, with the fifth state being questionable.

Broadening the total cropland analyses to all states, showed that June Agricultural Survey data could be carried forward to the September Agricultural Survey for the following twenty-two states: Arkansas, Delaware, Georgia, Idaho, Illinois, Indiana, Iowa, Louisiana, Maine, Michigan, Minnesota, Mississippi, Montana, Nebraska, North Dakota, Ohio, Pennsylvania, South Dakota, Utah, Virginia, Washington, and Wisconsin. However, for this to occur, the quality of June data needs to be improved.

One of the recommendations was to develop an analysis table in NASS' Interactive Data Analysis System (IDAS). This analysis table would indicate all of the operations with positive values for those items reported in June that were also sampled for September. The table would also display the operations' expansion weights for September, their reported values for these items in June, and the resulting indications derived by multiplying these two items. Statisticians would then be able to see the impact of carrying the June data forward and identify operations that may be problematic.

Previously Reported Data Usage in NASS Field Offices - 2009, Gottschall, C., Census and Survey Division, DCB-09-01, United States Department of Agriculture, National Agricultural Statistics Service.

Using previously reported data (PRD) in surveys has long been an intriguing topic in the survey community. A variety of survey organizations, including the USDA's, National Agricultural Statistics Service (NASS) have experimented with and/or currently use PRD in establishment surveys. At the present, NASS uses previously reported data (PRD) in variety of nationally administered surveys. The data are primarily used for edit checks in headquarters (HQ) developed computer assisted telephone instruments (CATI) and interactive editing systems, but are also used to reduce contacts or in lieu of asking certain questions.

Table 1: Previously Reported Data Research within the National Agricultural Statistics Service (Continued)

Continued…

NASS Field Offices (FOs) also reported using additional PRD for both pre-response prompts as well as post-response check/verification in select surveys. Some FOs also provided PRD as background information or interview preparation for enumerators. The most intense uses reported were on smaller, state-oriented surveys where variables were expected to be largely unchanged (fruit acres); there were multiple contacts in a cycle (potatoes); or where the data were used, not only for the current survey, but also as a source of control data verification and updates.

FOs also expressed interest in employing more PRD in a variety of situations. Most suggested using PRD for pre-response prompts (verification) or to skip questions entirely in follow-on surveys. FOs also acknowledged concerns or challenges in utilizing PRD, including confidentiality issues, possible differing usages across modes, and introducing potential biases into the response process.

Based on these suggestions for expanded PRD uses, an analysis was conducted into the amount of change actually occurring in static variables over a crop year. The results were mixed. There did appear to be a fair amount of change or difference in the variables tested from quarter to quarter. Additionally, there were a fairly limited number of complete, useable records for which data could be pulled forward and used as PRD. Finally, there were actually very few directly comparable variables that can utilize PRD in a given crop year.

RECOMMENDATIONS:

1. Utilizing PRD as a pre-response prompt must be consistent across modes. Currently, there are significant logistic/operational difficulties in printing PRD on paper questionnaires using externally printed instruments (QAS, livestock surveys, etc.). Furthermore, there may be a difference in mental response processes used to analyze a follow-up clarification question (e.g. a built in Computer Assisted Telephone Interview (CATI) edit check) as opposed to answering an open ended question or confirming/updating PRD pre-printed or asked as a pre-response prompt (Sudman, et al 1996; Kalton and Schuman, 1982). This concept needs to be thoroughly tested prior to a large scale rollout.

2. The literature is not as rich regarding possible differences between providing PRD in a Web survey as compared to a paper instrument. Most researchers agree that a Web instrument, all things equal, is largely equivalent to its corresponding paper questionnaire (Dillman and Smyth, 2007). This raises the question of how PRD would be used in a Web instrument:

a.) As a type of built in edit check. This becomes inconsistent with the paper version (which has no interactive ability), may prove frustrating to respondents, could possibly slow down the Web interaction, and could lead to more break-offs.

b.) Simply as a pre-response prompt for confirmation or clarification. Several questions arise (and, at least currently, tend to be minimally researched): how will respondents react to seeing their personal PRD on a Web instrument—is it a different emotion than seeing PRD on a paper survey; will respondents be more, less or equivalently likely (or less likely) to update data on the Web compared to paper; and finally, how are inconsistent data handled—will respondents expect automated summations, tabulations, etc. in a Web version and how will they react to discrepancies and/or lack of expected built-in operations?

3. For all surveys that have an Electronic Data Reporting (EDR) and CATI component, and the CATI instrument utilizes PRD for built in skips of certain static variables already reported in the current survey cycle, EDR should incorporate the same skips as CATI. This design has already been built into the Quarterly Labor Survey and additional implementation is slated for additional surveys, beginning with the Quarterly Agricultural Survey.

4. Using pre-response prompts may reduce cognitive burden and develop or improve some respondent goodwill. However, for many of the proposed variables it seems unlikely to reduce overall interview time.

Table 1: Previously Reported Data Research within the National Agricultural Statistics Service (Continued)

Continued …

5. Providing previously reported expected yields as a pre-response prompt and asking if anything had changed appears to expose the data collection to possible measurement error due to faulty cognitive heuristics (Kalton and Citro, 1993).

6. Detailed operation profiles could be developed for more surveys and all States using Data Warehouse current crop year information. Several FOs currently use prototypes that they have developed for use in their State's data collection. These profiles would be tailored to a specific survey and could be either electronic or printed on paper.

7. As already noted, a fair number of FOs utilize PRD for a variety of non-probability, specialty commodity surveys (potatoes, fruit, etc.). Given the structure and process of editing, summarizing, and estimating for these unique variables, this seems like an appropriate utilization of PRD and a possible area of expansion.

8. Finally, the effects of using PRD to skip questions should be back tested using historical datasets, such as from a year's worth of Quarterly Agricultural Surveys. This would provide conclusive evidence as to the possible value or shortcomings PRD. However, even if testing shows current data to be consistent with PRD, caution must be exercised in using PRD to skip follow-on questions, given the nature of crop and livestock growing cycles.

The Effect of Different Methods of Providing Previously Reported Data on Current Reports of Cattle on Feed Inventory - 1995, Stanley, J. S., Survey Management Division, DCB-94-01, United States Department of Agriculture, National Agricultural Statistics Service.

This study investigated if using PRD in CATI edits for the Cattle on Feed Survey would reduce respondent burden. CATI respondents in South Dakota's April 1993 Cattle on Feed (COF) Survey were randomly assigned to three groups. One group was provided with their previous COF inventory number when their current total inventory, minus placements, plus disappearances did not equal their previous number (this is the approach used on the current operational survey). A second group was asked the same questions as the first group (total inventory, followed by placements and marketings) but was provided with their previous inventory number before they provided any current information. The third group was also provided with their previous number up front but was next asked for placements and marketings followed by current total inventory. The number of times data were edited during the interview was the largest for the first group which received the previously reported data as an edit check. This group also had the highest number of edits to the data after the interview. Differing administrations of previous quarter data appeared to have no significant effect on the total inventory reported with both the mean inventory reported and the change in total inventory from the previous to current quarter being the same for all groups. The results indicated that using PRD for Cattle on Feed CATI edits reduced respondent burden and had a minor impact on the numbers reported. The study also found that enumerators did not like using PRD as an edit check since they felt uncomfortable telling respondents that they were "wrong" after the respondents provided their answers. Enumerators preferred to provide the respondents with the PRD up front.

Result: Several previously reported data items are used as edit checks for Cattle on Feed including capacity and inventory.

Evaluation of Using Previously Reported Data in the 1993 Agricultural Yield Surveys - 1994, Bailey, J. T., Research Division, SRB-94-02, United States Department of Agriculture, National Agricultural Statistics Service.

This study examined the use of PRD for winter wheat, corn, and soybeans in CATI edit checks for the 1993 Agricultural Yield Surveys. During the interviews, respondents' answers for planted acres and yield were compared with PRD obtained from previous months' Ag. Yield Surveys. Respondents were asked to verify acreage and yield responses that were outside the prescribed range from the PRD. The results concluded that PRD should be used for acreage edit checks in the Agricultural Yield Surveys.

Result: Previously reported data are used for acreage and yield as edit checks.

Table 1: Previously Reported Data Research within the National Agricultural Statistics Service (Continued)

An Examination of the Cognitive Process Involved in Grain Stocks and Storage Capacity Reporting - 1993, Stanley, J. S., Survey Management Division, DCB-93-02, United States Department of Agriculture, National Agricultural Statistics Service.

This study investigated how respondents arrived at their answers for reported corn and soybean stocks inventory and grain storage capacity during the Quarterly Agricultural Survey. The study used a sub-sample from Ohio's 1993 March QAS and PRD from the December 1992 QAS.

The results indicated that respondents were generally unable to re-report their last quarter's stocks inventory and storage capacity correctly and did not think that having their previously reported numbers would increase the ease of reporting or the accuracy of their current report. When provided with their PRD on storage capacity and asked about the accuracy of these numbers, very few respondent felt the numbers were exactly correct. However, the results also suggest that supplying PRD on storage capacity to respondents may improve the quality for this item.

Evaluation of Historical Data Use in the 1992 August Yield Survey - 1993, Bailey, J. T., Research Division, SRB-93-02, United States Department of Agriculture, National Agricultural Statistics Service.

This study examined the use of PRD for corn harvested acres and soybean planted acres in CATI edit checks for the 1992 Agricultural Yield Survey. During the interviews, respondents' answers for planted acres and yield were compared with PRD obtained from previous months' Ag. Yield Surveys. Respondents were asked to verify acreage and yield responses that were outside the prescribed range from the PRD.

The results showed that, in about half of the states, there was a statistically significant difference between the initial and final August responses (as modified from the PRD edits) for both corn harvested acres and soybean planted acres. Soybean planted acres were significantly different at the US level; however, corn harvested acres were not.

The Effect of Using Historical Data in CATI Grain Stocks Enumeration in the March 1988 Agricultural Survey -1992, Mergerson, J. W. and O'Connor, T., Research and Applications Division, SRB-92-01, United States Department of Agriculture, National Agricultural Statistics Service.

This 14-state study investigated whether using corn, soybean, and all wheat historical data in data collection edits for on-farm grain stocks and on-farm storage capacity would reduce respondent burden by minimizing the number of call backs to clarify questionable reports. The results showed that historical data may improve the quality of grain stocks data. After being prompted with the PRD edits, respondents tended to increase their original responses upward, resulting in survey indications being closer to the Agricultural Statistics Board balance sheet indications. However, some respondents did not like knowing that we have their historical data. These respondents were under the impression that all past questionnaires were shredded.

The Influence of Using Previous Survey Data in the 1986 April ISP Grain Stocks Survey - 1988, Pafford, B., Research and Applications Division, Research and Applications Division, SRB-88-01, United States Department of Agriculture, National Agricultural Statistics Service.

A research study was undertaken for the April 1986 Integrated Survey Program (ISP) Survey (now called the Quarterly Agricultural Survey Program) in Califomia, 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

Table 1: Previously Reported Data Research within the National Agricultural Statistics Service (Continued)

Continued …

to 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 com 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 com storage increased these may not have been real increases. That is, problems in getting the correct January com 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.

Response Errors in NASS Surveys: The Effect of Using Previous Survey Data in the 1985 California Fall Acreage and Production Survey – 1986, Pafford, B., Research and Applications Division, SF&SRB-86-99, United States Department of Agriculture, National Agricultural Statistics Service.

This study examined the effects of using planted acreage from the 1985 June Acreage Survey on responses to planted acreage on the 1985 Fall Acreage & Production Survey (CATI only in California). This study used four treatment groups: a.) no PRD, b.) enumerators had access to PRD to use as they liked. c.) PRD was worded directly in the questions d.) when PRD were present, “planted acre” questions were skipped and respondents were asked only harvested acre questions.

The author recommended cautious use of historical data until further studies can be done. The author did not recommend showing PRD on the CATI screen for the enumerator to decide whether and how to use the data. Respondents who were not presented with any PRD under reported their crop acreage. However, when the respondents were provided with PRD embedded in the questions there was considerable re-reporting of the same numbers.

1980 - The Effects of Omitting Acreage Questions and Modifying the Operation Description Section in Cattle Surveys (SRD Working Paper) - 1980, Nealon, J., Statistical Research Division, SRD Working Paper, United States Department of Agriculture, National Agricultural Statistics Service.

Results indicated that the removal of the "acres in operation" question did not significantly increase the refusal rate. The estimate of the proportion of zero cattle operations and the estimates of the four selected survey items were rarely significantly different between the operational and test questionnaires. Finally, the proportion of individual operations reporting partnership arrangements was not significantly different between the operation description sections. Therefore, the acreage questions can probably be removed from the operational questionnaire and the test version of the operation description section implemented without significantly affecting the survey estimates. This statement implies only that the estimates should not change significantly and not that the estimates are or are not accurate.

3. CURRENT USES OF PRD DURING DATA COLLECTION AT NASS

Table 2 summarizes NASS' most wide-spread uses of PRD during data collection. This table originated from Beckler and Farrar in a proposal to the NASS Program Planning Committee in March 2006. Details of the proposal are shown in Appendix A. Gottschall (2009) also discusses several of these uses.

Table 2: Current Uses of Previously Reported Data during Data Collection within NASS.

Mode	Usage Category	Survey	Description
CATI	In lieu of asking questions	Agricultural Labor, October Quarter	For those done in CATI, if a completed report was received in the July quarter then the following questions are skipped (July data are brought forward): ▪ Screening questions ▪ Gross Value of Sales (or points) ▪ Farm Type ▪ Peak Hired Workers ▪ Reporting Unit ▪ Partner information ▪ Additional operation information
CATI	In lieu of asking questions	Agricultural Labor, January quarter	For those done in CATI, if a completed report was received in the July or October quarters the following questions are skipped (July/October data are brought forward): ▪ Screening questions ▪ Reporting Unit ▪ Partner information ▪ Additional operation information
CATI	In lieu of asking questions	Agricultural Labor, April quarter	For those done in CATI, if a completed report was received in the July quarter the following questions are skipped (July data are brought forward): ▪ Screening questions ▪ Reporting Unit ▪ Partner information ▪ Additional operation information For those done in CATI, if a completed report was received in the January quarter the following questions are skipped (January data are brought forward): ▪ Screening questions ▪ Gross Value of Sales (or points) ▪ Farm Type ▪ Peak Hired Workers ▪ Reporting Unit ▪ Partner information ▪ Additional operation information
CATI	In lieu of asking questions	Agricultural Yield Survey	Commodity questions are skipped in the CATI instrument if the respondent did not indicate those crops were planted on the Crops/Stocks Survey.
CATI	In lieu of asking questions	Crops/Stocks Surveys, September, December, and March quarters	For those done in CATI, if a completed report was received in the June quarter the following questions are skipped (June data are brought forward): ▪ Screening questions ▪ Reporting Unit ▪ Partner information ▪ Acres Operated

Table 2: Current Uses of Previously Reported Data during Data Collection within NASS. (Continued)

Mode	Usage Category	Survey	Description
Continued from previous page.			Note: If the operation is a refusal or inaccessible in the follow-on quarters, Cropland, All Land Operated, and Total Storage Capacity are pulled from the sample master.
CATI	In lieu of a contact	Crops/Stocks Survey, September quarter	Operations in the September Crops/Stocks Survey sample are not contacted if they were also in the June Crops/Stocks Survey and reported no items of interest (for the September Survey) in June.
CATI	CATI edits	Milk Production Survey, Cattle on Feed, Bee and Honey, Ag Labor, Maple Syrup, Ag Yield, Quarterly Hogs, Layers and Eggs	Various previously reported data items are used as edit checks and consistency edits for the respondent, interviewer and during interactive editing.
EDR[1]	PRD supplied in current question	Crop Progress & Condition	Respondents see their previous week's response before answering current week conditions.
Paper	PRD supplied in current question	Crops/Stocks Surveys, Cattle on Feed Survey, Chickens Survey	Some FOs provide enumerators with PRD either printed on the label, hand written elsewhere on questionnaire, or on supplemental sheets of paper. It should be noted that most, if not all, of these practices are likely in violation of (Policy and Standards Memo – 47: Data Collection Policy and Standards).
Paper	PRD supplied in current question	Cold Storage Biennial Capacity Survey	Each facility's capacity is preprinted on the questionnaires; this capacity came from a previous Capacity Survey.
CATI and Paper	PRD supplied to enumerator	Cattle on Feed Survey	Enumerators are provided with the previous month's reported cattle on feed inventory.
CATI	In lieu of asking questions	Monthly Chickens and Eggs	The Chickens and Eggs questionnaire contains questions pertaining to table egg flocks, broiler-type hatchery supply flocks, and egg-type hatchery supply flock. Operations do not tend to switch between types of birds raised; therefore, historical reporting is used during interviews to only ask questions relating to the given operation's bird type.
CATI	In lieu of a contact	Monthly Chickens and Eggs	If the operation reports as out-of-scope for the current survey and does not plan on being in scope (greater than or equal to 30,000 table egg layers), operation is not contacted for future months, instead is included in the less than 30,000 estimate.

[1] EDR: Electronic Data Reporting or web base questionnaire.

4. FUTURE USES OF PRD DURING DATA COLLECTION

Table 3 summarizes additional potential future uses of PRD during data collection by NASS. Specific implementations of any method would require empirical evidence to justify the benefit. This table was based on the 2006 presentation to the Program Planning Council.

Table 3: PRD Usage During Data Collection

Usage Category	Description	Potential Data Quality Effect[1/]	Potential Respondent Burden Effects[1/]
EDR edits[2/]	Reported values are compared with PRD; if the currently reported value is deemed inconsistent with the PRD, the respondent is asked to verify the current answer or explain the situation.	Improve	Lessen
In lieu of a contact[3/]	PRD are used to avoid contacting an operation. This may take two forms: (a) the PRD are brought forward and used to help model data for the current survey, or (b) the PRD may be used to determine that the operation is out-of-scope for the current survey (i.e., a survey zero).	Worsen	Lessen
In lieu of asking questions[3/]	PRD are used to avoid one or more questions. This may take two forms: (a) the PRD are brought forward and used for the current survey, or (b) the PRD may be used to determine that the operation has no pertinent data for one or more questions in the current survey (i.e., used in branching/skip patterns).	Worsen	Lessen
PRD supplied in current question (or to enumerator)	PRD are provided to respondents to assist with answering current questions. This may be: (a) to verify a previously reported answer, (b) to make providing a current answer easier, or (c) as a starting point for a balance sheet.	Unknown	Worsen or Lessen (Situational Dependent)
Tailoring data collection	PRD are used to improve the data collection process. Includes such things as altering question includes/excludes or providing more/fewer prompts to the enumerator.	Unknown	Lessen

1/ The potential effects are based on research conducted by NASS.

2/ EDR edits would reduce respondent burden if they prevented a call-back. These edits, however, may also increase respondent burden if resolving errors/warnings lengthen the interview time.

3/ H. Tran and M Gerling 2009 and 2010, found that bringing forward (cropland, total land, land rented from, land rented to, and/or storage capacity) from the June Agricultural Survey to September's Agricultural Survey can have a negative impact on the data.

5. CONSIDERATIONS FOR INCREASING THE USE OF PRD IN NASS SURVEYS

Increasing the use of PRD, requires NASS policy to be re-evaluated and current procedures revised. To expand the use of PRD, the following issues need to be addressed:

a.) While providing PRD to respondents may lessen response burden for some respondents, it will decrease the ability to detect true change or detect reporting errors in the items targeted.

   1. NASS would also have to decide whether PRD will be shared with respondents other than the one who reported the PRD. This requires revision to NASS' confidentiality of data clause. If not, procedures to enable identification of the person reporting an item to be used as PRD would have to be developed.

b.) The availability of current PRD to be used in survey data collection is limited. NASS' sampling approach is to minimize the number of times any particular agricultural operation is sampled across surveys. This methodology is in direct opposition to having an abundant and current pool of PRD to utilize. Also, Field Office's special handling of large operations to collect data only 1-2 times a year, significantly limits the amount of PRD available for these operations.

c.) Since PRD is not available for a given item for all operations, survey questions which incorporate PRD as well as questions that do not have PRD would have to be developed and used as appropriate. This would require developing and conducting usability testing on the appropriate way to incorporate PRD across each survey mode.

d.) NASS currently retains the data passed on to the summary and does not keep the originally reported value when data are edited. The final values retained are a combination of respondent, edited, and imputed. Hence, NASS would have to decide what data would be used as PRD and if reported data is defined as PRD then NASS would have to start capturing and storing reported data before it is edited. In addition, NASS would have to decide on whether respondent PRD or edited PRD would be used in edit checks.

e.) Procedures for capturing and retaining any PRD that were corrected by respondents would have to be developed and incorporated into NASS' systems.

f.) As documented in H. Tran, M. Gerling (2009, 2010) and C. Gottschall (2009), the quality of the data collected needs to be improved. Adding additional edit checks at the time the data are being collected and at the time of editing would need to be enhanced.

6. CONCLUSION

The National Agricultural Statistics Service has done substantial research and examination into using previously reported data during data collection. In general, using accurate, previously reported data has the potential to reduce respondent burden. However, using previously reported data also has the potential to adversely affect data quality. As discussed in Section 5, substantial altering of NASS current processes, policies and systems are needed to fully implement the use of previously reported data. This would not only be a paradigm shift for NASS but also requires staffing resources to complete.

7. RECOMMENDATION

The National Agricultural Statistics Service's Strategic Planning Council decide whether the agency is willing to significantly revise current NASS policies, processes, systems, and operations to increase the use of PRD to reduce response burden for a subset of respondents, while at the same time accepting a potential decrease in data quality.

Appendix A

Previously Reported Data (PRD) Action Plan
March 2006 Program Planning Council Meeting

Previously Reported Data (PRD) Action Plan

March 2006 Program Planning Council Meeting

1. Background

The use of previously reported data (PRD) during data collection may improve data quality as well as reduce respondent burden. However, used improperly, PRD may also increase measurement error, leading to biased survey results and even increase respondent burden. This document summarizes NASS' past research and current operational uses of PRD during data collection. It also presents a plan to expand the use of PRD. This plan is based on the premise that the primary motivation to expand the use of PRD during data collection is to reduce respondent burden (as measured by our Joint Burden Indicators). Throughout this document "PRD" refers only to data previously provided to NASS; this definition is consistent with PSM-47. The use of administrative data – those obtained from non-NASS sources – is not addressed.

2. Possible PRD Usage During Data Collection

Table 1 briefly summarizes the various ways PRD may be used during data collection. The *Potential Benefit* column identifies whether each method would primarily be used to improve data quality or reduce respondent burden. Specific implementations of any method would require empirical evidence to justify the benefit.

Table 1: PRD Usage During Data Collection

Usage Category	Description	Potential Benefit	
		Data Quality	Respondent Burden
CATI/EDR edits	Reported values are compared with PRD; if the currently reported value is deemed inconsistent with the PRD, the respondent is asked to verify the current answer or explain the situation.	●	● 1/
In lieu of a contact	PRD are used to avoid contacting an operation. This may take two forms: (a) the PRD are brought forward and used to help model data for the current survey, or (b) the PRD may be used to determine that the operation is out-of-scope for the current survey (i.e., a survey zero).		●
In lieu of asking questions	PRD are used to avoid one or more questions. This may take two forms: (a) the PRD are brought forward and used for the current survey, or (b) the PRD may be used to determine that the operation has no pertinent data for one or more questions in the current survey (i.e., used in branching/skip patterns).		●
PRD supplied in current question (or to enumerator)	PRD are provided to respondents to assist with answering current question(s). This may be: (a) to verify a previously reported answer, (b) to make providing a current answer easier, or (c) as a starting point for a balance sheet.	●	●
Tailoring data collection	PRD are used to improve the data collection process. This may include such things as altering question includes/excludes or providing more/fewer prompts to the enumerator.	●	●

1/ CATI/EDR edits reduce respondent burden if they prevent a call-back. CATI/EDR edits may also increase respondent burden if resolving errors/warnings lengthen the interview time.

3. Past NASS PRD Research

Table 2 summarizes NASS' research reports on using PRD. (This table has been excluded from the body of Appendix A since this is similar to Section 1 of this report).

4. Current NASS Uses of PRD During Data Collection

Table 3 provides examples of NASS' most wide-spread uses of PRD during data collection. The *Operations Affected* column shows the number of operations for which PRD was used in the manner described for the indicated survey date.

Table 3: Current NASS PRD Usage During Data Collection

Mode	Usage Category	Survey	Description	Operations Affected [1/]
CATI	In lieu of asking questions	Agricultural Labor, October quarter	For those done in CATI, if a completed report was received in the July quarter the following questions are skipped (July data are brought forward): ▪ Screening questions ▪ Gross Value of Sales (or points) ▪ Farm Type ▪ Peak Hired Workers ▪ Reporting Unit ▪ Partner information ▪ Additional operation information	3,011 (Oct. 2005)
CATI	In lieu of asking questions	Agricultural Labor, January quarter	For those done in CATI, if a completed report was received in the July or October quarters the following questions are skipped (July/October data are brought forward): ▪ Screening questions ▪ Reporting Unit ▪ Partner information ▪ Additional operation information	2,608 (Jan. 2004)
CATI	In lieu of asking questions	Agricultural Labor, April quarter	For those done in CATI, if a completed report was received in the July quarter the following questions are skipped (July data are brought forward): ▪ Screening questions ▪ Reporting Unit ▪ Partner information ▪ Additional operation information For those done in CATI, if a completed report was received in the January quarter the following questions are skipped (January data are brought forward): ▪ Screening questions ▪ Gross Value of Sales (or points) ▪ Farm Type ▪ Peak Hired Workers ▪ Reporting Unit ▪ Partner information ▪ Additional operation information	4,945 (Apr. 2005)

Table 3: Current NASS PRD Usage During Data Collection (Continued)

Mode	Usage Category	Survey	Description	Operations Affected [1/]
CATI	In lieu of a contact	Crops/Stocks Survey, September quarter	Operations in the September Crops/Stocks Survey sample are not contacted if they were also in the June Crops/Stocks Survey and reported no items of interest (for the September Survey) in June.	8,365 (Sept. 2005)
CATI	In lieu of asking questions	Agricultural Yield Survey	Commodity questions are skipped in the CATI instrument if the respondent did not indicate those crops were planted on the Crops/Stocks Survey.	24,869 (Aug. 2005, largest AY sample)
CATI	CATI edits	Milk Production Survey, Cattle on Feed, Bee and Honey, Ag Labor, Maple Syrup, Ag Yield, Quarterly Hogs, Layers and Eggs	Various previously reported data items are used as edit checks and consistently edits for the respondent, interviewer and during interactive editing.	N/A
EDR	PRD supplied in current question	Crop Progress & Condition	Respondents see their previous week's response before answering current week conditions.	N/A
Paper	PRD supplied in current question	Crops/Stocks Surveys, Cattle on Feed Survey, Chickens Survey	Some FOs provide enumerators with PRD either printed on the label, hand written elsewhere on questionnaire, or on supplemental sheets of paper. It should be noted that most, if not all, of these practices are likely in violation of PSM-47.	N/A
Paper	PRD supplied in current question	Cold Storage Biennial Capacity Survey	Each facility's capacity is preprinted on the questionnaires; this capacity came from a previous Capacity Survey.	N/A
CATI and Paper	PRD supplied to enumerator	Cattle on Feed Survey	Enumerators are provided with the previous month's reported cattle on feed inventory.	N/A
CATI	In lieu of asking questions	Monthly Chickens and Eggs	The Chickens and Eggs questionnaire contains questions pertaining to table egg flocks, broiler-type hatchery supply flocks, and egg-type hatchery supply flock. Operations do not tend to switch between types of birds raised; therefore, historic reporting is used during interviews to only ask questions relating to the given operation's bird type.	N/A
CATI	In lieu of a contact	Monthly Chickens and Eggs	If the operation reports as out-of-scope for the current survey and does not plan on being in scope (greater than or equal to 30,000 table egg layers), operation is not contacted for future months, instead is included in the less than 30,000 estimate.	N/A

1/ Reflects the number of operations to which PRD was used in the manner described for the indicated survey date.

Previously Reported Data (PRD) Action Plan

5. Availability of PRD

NASS' data warehouse may eventually contain all data that are used as PRD. While the data warehouse currently contains literally billions of rows of data, for a variety of reasons, much of these data may not be useful as PRD. Perhaps the foremost reason why warehouse data may be inappropriate for PRD is the lack of sufficiently current data for operations in current survey samples.

Determining exactly how much *usable* PRD data are available is difficult; however, the following two examples offer some insight into how much might be available.

Example 1

This example looks at how many list frame records were contacted for major sample surveys between January 2000 and December 2003. Table 4 was recreated from Suzette Qualey's 2005 research report, *Accumulated Respondent Burden and Response Rates in Surveys by USDA's National Agricultural Statistics Service*. The list frame *Blue Books* show that there were approximately 1.8 million active farms on the frame during the period summarized in Table 4. Therefore, excluding data from "small surveys", approximately ⅓ of the active list frame records could potentially have reasonably current PRD (579,531 out of 1.8 million). In fact, this is generous because Table 4 does not consider nonresponse.

Table 4: Frequency of Operations by Number of Major Survey Contacts 1/ (Jan. 2000 – Dec. 2003)

Number of Surveys	Frequency of Operations	Percent of Operations
1 – 5	467,662	80.70
6 – 10	72,642	12.53
11 – 15	24,773	4.27
16 – 20	8,115	1.40
21 – 25	2,118	0.37
26 – 30	826	0.14
31 – 35	491	0.08
36 – 40	439	0.08
41 – 45	460	0.08
46 – 50	211	0.04
51 – 55	190	0.03
56 – 60	932	0.16
61 – 65	374	0.06
66 – 70	161	0.03
71 – 75	68	0.01
76 – 80	36	0.01
81 – 85	13	< 0.01
86 – 90	9	< 0.01
91 – 95	4	< 0.01
96 – 100	4	< 0.01
101 – 105	3	< 0.01
Total	579,531	100.00

1/ Excludes the 2002 Census of Agriculture

Example 2

Another example that shows the potential availability of PRD is shown in Table 5. This table shows the amount of sample overlap between each quarter's Crops/Stocks Survey and any previous quarter for classify year 2005. (The amount of sample overlap is controllable.) While this table does not consider nonresponse, it is clear that for the Crops/Stocks Survey, the majority of operations sampled may have PRD from a previous quarter for the same crop year (considering only winter wheat for March).

Table 5: Classify Year 2005 Crops/Stocks Survey Sample Overlap

Quarter	Sample Size	Samples in Any Previous Quarter[1/]	
		Count	Percent
June 2005	74,026 [2/]	–	–
September 2005	60,098	39,949	66.5
December 2005	79,191	61,085	77.2
March 2006	79,565 [2/]	79,565	100

1/ In Classify Year 2005 Crops/Stocks Survey quarters.
2/ Includes Agricultural Yield samples.

6. PRD Expansion Plan

The following six-part plan is offered to expand the use of previously reported data. In accordance with PSM-47, the effects of PRD usage (during data collection) on survey indications, respondent burden, and response rates must be evaluated prior to full program implementation. The following plan provides for this evaluation, when appropriate, as well as an implementation strategy.

Part 1: Determine Specific PRD Goals and Definitions

As stated earlier, it is assumed that the primary motivation to expand the use of PRD during data collection is to reduce respondent burden (as measured by our Joint Burden Indicators). However, there are at least two meanings of "respondent burden reduction". One meaning is to reduce, on average, the amount of NASS-imposed respondent burden for each individual operation. Basically, this amounts to some combination of fewer surveys, fewer (or perhaps easier) questions, or fewer contacts for each individual operation. The second meaning of respondent burden reduction is to reduce the total respondent burden for the population.

By increasing the amount of sample overlap between surveys, more PRD would likely be available for more individual operations. This would allow for the maximum use of data collections strategies employing PRD, and would thus reduce the overall Joint Burden Indicator for "amount of time" for the population. Greater overlap between samples could also reduce some sample sizes by benefiting from stronger c/c indications. Obviously, though, increasing sample overlap would increase the burden to some extent on specific individual operations.

Hence, NASS should consider what the overall goal is for PRD: (1) to reduce individual respondent burden, or (2) to reduce overall respondent burden for the population. PRD strategies employed may differ depending on the goal.

Another issue that should be addressed is to define exactly what data may be considered *usable* PRD. PSM-47 provides some guidance on using PRD, but it does not provide enough detail. Specific questions concerning what may be considered *usable* PRD include:

- Should only current crop year information be considered for PRD? (Example: should three year old corn planted acres be used as PRD?)
- Should PRD be limited to items that are "static"? If so, how is "static" defined? Is the definition of "static" different for different items?
- Are any categories of data off limits? (Example: should price data ever be used as PRD?) definition of "static" different for different items?

- What items within what surveys should be considered for PRD usage? What research is needed?

Part 2: Implement Existing CATI Uses of PRD Into EDR

NASS' electronic data reporting (EDR) system allows conditional question routing and the use of data from an external file (that was loaded with the EDR sample). Since the capability exists for EDR to incorporate all existing usages of PRD, and since these usages are already being used for CATI, the uses of PRD as documented in Table 3 will be incorporated into the EDR instruments by the next practicable instance of each survey.

Part 3: Reusing "Planted Acres"

Research Step

Identify all surveys that ask "Planted Acres" for any commodity. Perform "simulation studies" with the past 2-3 years of the September and December Crops/Stocks Surveys where direct expansions are compared from the individual Crops/Stocks Surveys against direct expansions obtained from substituting previously reported commodity-level acres planted data from the same crop year (collected earlier than the Crops/Stocks Survey quarter being considered).

Example: For the 2005 September Crops/Stocks Survey obtain the direct expansions for each commodity's planted acres using the reported survey data. Next, use the same 2005 September Crops/Stocks direct expansion weights, calculate new direct expansions where September data are substituted with 2005 June Crops/Stocks Survey reported planted acres for all records that were in both the June and September quarters (September data would be used for all operations not in June). The two sets of direct expansions would be compared to determine the effect of using PRD in the 2005 September Crops/Stocks Survey.

Impact Step

Assuming the direct expansion comparisons in the *Research Step* yield acceptable PRD effects, the impact on respondents will be estimated. This will involve estimating how many total burden hours are saved, based on how many questions are "skipped" due to using PRD.

Implementation Step

Assuming the direct expansion comparisons in the *Research Step* yield acceptable PRD effects, research will be conducted to determine the most effective way to utilize the planted acres PRD. The following approaches will be considered:

1. Skip the "planted" questions for each commodity and simply ask for the number of acres harvested. This approach has three advantages: (1) it maximizes respondent burden savings, (2) it avoids the issue of ensuring that the PRD are revealed only to appropriate respondents, and (3) it avoids the issue of ensuring that only actually reported data are provided to the respondent. However, the disadvantage of this approach is that harvested acres may exceed the PRD planted acres. When/if this occurs we could modify our edits to address such situations.
2. Provide the PRD planted acres in the "harvested" question and allow the enumerator to change them only if the respondent indicates the PRD are incorrect. (Example: Our records indicate you have 1,000 acres of oats; how many acres of oats have you harvested or intend to harvest? Only if the respondent indicates the 1,000 acres of oats is incorrect would the respondent be asked to provide a revised figure.) The advantage of this approach is greater savings in respondent burden; the disadvantage is this approach may miss some changes to PRD (however, the "research step" presumably showed such changes did not dramatically affect survey indications anyway).
3. Provide the PRD planted acres and ask the respondent if they are correct; if incorrect, the respondent would be asked to provide revised values otherwise, the respondent would be asked for acres harvested.

Once the specific approach to using the planted acres PRD is determined, PRD will be used for planted acres on the next practicable instance of the September and December Crops/Stocks Surveys.

Part 4: Capturing Reported, Edited, Imputed, and Final Data

According to PSM-47, only reported (i.e., keyed) data may be provided back to respondents. Often, data are hand edited by SSO staff before they are keyed. Some mechanism needs to be in place to distinguish data that came directly from a respondent from data that are estimated, changed or removed during a pre-keying SSO hand editing operation. Such a mechanism is currently not in place. In order to ensure that all surveys are capable of supplying PRD that may be provided back to respondents, steps will be taken to ensure that all surveys (and censuses) loaded to the data warehouse include at least the following data types: data as it is provided by the respondent, reported (keyed), imputed, edited, and final. This will involve modifications to Blaise CATI/IE instruments and to SPS for each survey. These modifications are anticipated to require a fairly large amount of time. Hence, surveys will be "retrofitted" as time permits.

Part 5: Electronic Data Reporting Risk Assessment

A risk assessment will be done to ensure that NASS meets Departmental cyber security standards when providing PRD to respondents in EDR.

Part 6: Research 2012 Census of Agriculture

The EDR instrument for the 2012 Census of Agriculture may provide PRD for each commodity's 2012 harvested acres to Web respondents. This could be simulated in a similar way to that described in Part 3 of Section 6 of this document using 2007 Census of Agriculture data. Harvested acres data from the most recent survey year would be substituted for the harvested acres data collected on the Census for a sample of Census records. Direct expansions would then be calculated and compared for the "real" Census data and the "simulated" Census data. This simulation study may have to be limited in scope, however, because the complex Census calibration and estimation processes would have to be used on both samples.