



United States
Department of
Agriculture



National
Agricultural
Statistics
Service

Census and Survey Division
Washington DC 20250

CSD Staff Report
Number DCB-09-01

June 2009

Previously Reported Data Usage in NASS Field Offices

Chris Gottschall

This report was prepared for limited distribution to the research community outside the United States Department of Agriculture. The views expressed herein are not necessarily those of NASS or USDA.

EXECUTIVE SUMMARY

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.

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 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:
 - 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.
 - 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 is 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.
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.

Previously Reported Data Usage in NASS Field Offices

Chris Gottschall¹

Abstract

This paper briefly examines the literature and prior research regarding previously reported data (PRD) in survey data collection. It then summarizes the USDA's National Agricultural Statistics Service (NASS) use of PRD at a national level. The paper also presents, in detail, current uses of PRD employed by individual NASS Field Offices (FOs) in their data collection processes. Based on existing literature, past research, NASS national level PRD usage, and FO uses and suggestions, several recommendations are developed for further research, expanded applications, and possible limitations of utilizing PRD in NASS data collection strategies. An appendix details a preliminary test of PRD in a data collection cycle.

Key Words: Agricultural Surveys, Previously Reported Data

1. INTRODUCTION

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. Social researchers have also employed dependent interviewing in an effort to utilize PRD in household or individual surveys (Mathiowetz and McGonagle, 2000). While the topic is of keen interest, research into utilization of PRD has been somewhat sporadic. Accordingly, the recommendations derived from the various experiments have typically been calls for more research.

There are several main reasons proffered by survey organizations for utilizing PRD. Reducing respondent burden is often a goal of incorporating PRD into the data collection process. By providing PRD to respondents, they can quickly confirm their previous responses. Along with reducing burden, PRD can hopefully reduce measurement error and/or improve data quality as respondents update or correct (as necessary) what they previously reported (Holmberg, 2004). PRD can also provide an anchor point for the current information request, eliminating seam effects (the difference in reported values for the same time period, collected on two different surveys) or avoiding possible spurious reports of change (Mathiowetz and McGonagle, 2000).

However, there are also concerns about extensive integration of PRD in data collection. A chief concern is respondents' utilizing the least cognitively burdensome heuristic in arriving at an answer ("satisficing" as coined by Krosnick (1991)) and simply agreeing with the PRD as presented to them (Stanley and Safer, 1994) There also exists the possibility of accidental

¹ Chris Gottschall is an Agricultural Statistician with the National Agricultural Statistics Service, Census and Survey Division, Data Collection Branch, 1400 Independence Avenue, SW, Washington, DC 20250.

disclosure of PRD to a respondent other than the previous data provider; providing back incorrect or edited/estimated data; or the instrument containing PRD simply being lost in the mail or compromised by circumstances outside the organizations control (Holmberg, 2004).

Finally, there may be very real costs or operational difficulties in incorporating PRD into certain modes (Holmberg, 2004). While it may be a “relatively” simple process to access a database and incorporate PRD into a computer assisted telephone interview (CATI) or Web survey instrument, it may be difficult or impossible to merge and print this data into a pre-printed questionnaire before mailing. If, due to these difficulties, PRD is utilized in one mode but not another, it leads to concerns about the standardization of the survey experience (and possible measurement error) across different modes (Biemer and Lyberg, 2003).

1.1 PRD in NASS

The USDA, NASS has also conducted internal research into utilizing PRD for data collection. Research has included:

- Cattle on Feed Surveys (Stanley, 1994): PRD are now utilized as edit checks for capacity and inventory.
- Agricultural Yield Surveys (Bailey, 1994): As a result, PRD are now utilized as edit checks for acreage and yield.
- Quarterly Agricultural Surveys (Stanley, 1993): Tested utilizing PRD for grain storage capacity and grain stocks reported in a prior survey.

As a result of this accumulated agency information, PRD are utilized at the national level in a variety of capacities, including (from Farrar and Beckler, 2006):

- Agricultural Labor follow-ons in CATI (in lieu of asking questions)
- Quarterly Agricultural Survey (QAS) follow-ons in CATI (as a post-response edit check, or in lieu of both asking questions or of a contact)
- Agricultural Yield Survey in CATI (in lieu of asking questions)
- Monthly Chicken and Eggs (in lieu of both asking questions or of a contact)
- A wide variety of major NASS surveys (at least 8) with HQ-developed CATI instruments use PRD for built-in edit checks
- Operation crop profiles for a variety of chemical surveys
- Crop Progress and Condition System provides crop weather web respondents with previously reported conditions

Additionally, some NASS Field Offices (FOs) use PRD specifically tailored to their individual data collection needs. Some of these applications may fall outside of the NASS data collection parameters as outlined in Policy and Standards Memorandum (PSM) 47, Policy on Data Collection. Upon review of the current research, like most of the establishment survey community, there was a wide range of utilization and attitudes towards PRD within NASS. FO comments in response to the data request ranged from:

“Our primary concern is that the agency will use methodological issues to keep FO's from using PRD. We need to quit worrying so much about the effects of PRD on quality. Quality is important but we need to understand that tradeoffs in quality must be made if we want to help ease reporting burden.”

to

“Whatever tools we are provided need to focus on obtaining correct responses, not necessarily making it easy on the survey stat, enumerator, or respondent.”

The current research focuses on developing a comprehensive list of these FO uses of PRD, summarizing the various aspects and procedures utilized by FOs to incorporate PRD into their data collection, evaluating the usages, and, finally, arriving at recommendations for implementation or additional analysis.

2. METHODS

On February 15, 2008, a request regarding current (or recent) uses of PRD was sent via email (Appendix B) to all the NASS FOs. This email contained a brief overview of the history and what was currently being researched regarding PRD uses in NASS; a summary of current Headquarters (HQ) PRD uses; and an Excel spreadsheet containing examples of FO uses of PRD, a space to summarize current (recent) FO uses of PRD, and FO proposed uses for PRD. All FOs were asked to respond, even if they were not using PRD in their data collection (other than the HQ PRD uses). As the information was received from individual FOs, the spreadsheets were then compiled into a single spreadsheet for summarization and analysis.

2.1 Information Collected

The information requested from the individual FOs was intended to provide HQ with detailed specifics regarding: 1) the extent of PRD usage in the FOs, and, 2) the procedures implemented by the FOs to handle the data and incorporate it into their data collection processes. Specifically, the inquiry requested information about ten aspects of PRD utilization:

- Survey for which the PRD were used
- Type of use (e.g. pre-response prompt, post-response check, route questions (skip), etc.)
- Specific item(s) on the survey (e.g. planted acres, storage capacity, inventory, etc.)
- In what situation(s) are the PRD used?
- Are the PRD used within this survey only (including follow-ons) or are they used in other surveys?
- How are enumerators trained to use the PRD?
- In which data collection mode(s) are the PRD used?
- How are edited data kept separate from reported data?
- How are the PRD stored and brought into the questionnaire and/or instruments?

- How well does this work?

Additionally, FOs were asked about proposed uses for PRD. This included questions about:

- For what surveys and uses should NASS consider employing PRD?
- What tools are needed to improve our ability to use PRD consistently and effectively?
- Other concerns (about PRD)?

3. RESULTS

Given the somewhat anecdotal nature of the information collected, the summarization presented here focuses on providing a thorough synopsis of the most common FO uses of PRD along with analysis of the procedures used by the FOs to incorporate PRD into their data collection procedures. As a result, and in an effort to best summarize the wide variety of information received in a relatively concise manner, the following presents the data in more of a narrative format than a typical research report.

Excluding the already noted HQ national applications, nearly two-thirds of NASS FOs reported using PRD in some form for at least one survey. FOs reported utilizing PRD on a wide variety of surveys ranging from the Quarterly Agricultural Survey to various specialty crop surveys to State County Estimates Surveys. PRD usages ranged from a limited number of items on only one survey (several FOs) to a wide variety of PRD for a substantial number of surveys (again, several FOs) with a whole host of variations of uses between the two extremes.

Generally, the FOs used PRD in telephone or personal interviews. A more limited number used pre-printed PRD on mail questionnaires. It appears that there may be some hesitancy in using pre-printed data—whether from a (very legitimate) concern about confidentiality, given minimal control over what happens to the questionnaire in the mail or who opens the instrument upon receipt; or that there are fairly difficult obstacles to implementing pre-printing on a large scale (Holmberg, 2004).

FOs used PRD in the collection of a wide variety of survey variables. Typically, they were fairly static items (total land, cropland, and various capacities) or items that become established and remain generally fixed over a season or cycle (crop acres, previous month's inventories). Most often, the PRD was used as a pre-response prompt/verification ("Last time, you reported 10,000 bushels of grain storage capacity. Is that still correct?"); or as a post-response check, similar to a CATI built-in edit check but on a non-CATI telephone or personal interview (respondent reports 1,000 bushels of capacity and interviewer follows with "Our records show 10,000 bushels of storage capacity for this operation. Was that incorrect?").

Procedures for integrating the PRD into the data collection system ranged from hand-copying a select number of items into questionnaires (very small sample sizes); printing items of interest values on special labels or on comment sheets; SAS programs that pulled PRD from historical datasets and created output that could be read into Survey Management System (SMS);

importing data from the Data Warehouse into SMS for FO developed CATI instruments; and various derivations of each across the spectrum of these procedures.

Table 1 provides a summary of the FO uses of PRD. Surveys are ranked according to the frequency of reported uses: Quarterly Agricultural Surveys were the most commonly reported FO utilization of PRD, fruit surveys were the second most common, etc. The table also summarizes how FOs reported using the PRD and for which variable on the respective survey.

TABLE 1: Previously Reported Data: Surveys, Uses and Items¹

Rank (frequency)	Survey	Most Common Uses	Most Frequent Items
1.	Quarterly Ag Survey	Pre-response prompt Post-response check	Total Land, Cropland, Grain Storage Capacity
2.	Fruit Surveys	Pre-response prompt Post-response check	Bearing Acres, Production, Varieties
3.	Potatoes	Pre-response prompt	Acreage, Yield, Stocks, Sales
	Grain Prices/Off-Farm Grain Stocks	Post-response check	Storage Capacity
	Cattle on Feed	Pre-response prompt, Post-response check	Capacity, Previous Inventory, Placements
6.	County Estimates Surveys	Post-response check	Total Land, Cropland, Crop Acres, Grain Storage Capacity, Livestock Items
	Hogs	Pre-response prompt, Post-response check	Inventory Items
8.	Cattle	Background information	Cattle inventory items of interest
	Vegetables	CATI edit checks, Background information	Planted Acres, Value of Sales
10.	Labor	Post-response check, Background information	Labor Count, Operation Information
	Dairy	Post-response check	January 1 Cows, Inventory Items
	Sheep	Background information	Sheep inventory items of interest
	Dry Beans	Post-response check	Planted Acres by Class

1: In addition to these, FOs (one for each survey) also reported using PRD on: Agriculture Resource Management Survey (ARMS), Garden Bean Seed Dealer Survey, Cotton Ginners, Pecans, June Area, Annual Manufactured Dairy, Prices Received by Farmers for Rice, Mink, Tobacco Inquiry, Maple Syrup Production, Bee and Honey, and Ag Yield.

3.1 Most Common Uses of PRD in FOs

This section provides additional detail on the surveys most frequently reported by FOs in which they use PRD. All of these reported uses are in addition to the HQ provided, national level applications already detailed in the introduction section.

Quarterly Agricultural Survey: Ten FOs used PRD in their Quarterly Agricultural Survey (QAS) data collection. Uses included pre-response prompts and post-response checks. Variables included total land (owned, rented from/to), cropland, crop acres planted and storage capacity. FOs reported data were pulled from Data Warehouse into SMS and then printed on labels or comment/call sheets. PRD were only used for personal interviews or telephone interviews completed on paper. No FOs pre-printed information on mail questionnaires. Since data were pulled from Data Warehouse, there was no way for FOs to keep edited data separate from reported data. Several FOs using the PRD as a post-response check commented that providing this information to enumerators was similar to the built-in CATI edit checks. One FO provided a wide assortment of Data Warehouse PRD via a special label to be used as background information for field enumerators prior to the interview.

Fruit Surveys: The second most frequent use of PRD among FOs was on fruit surveys. Data were most commonly used as a post-response check, although in a few cases it was used as a pre-response prompt. Typical variables were bearing acres/trees, yield or total production and varieties (by type of fruit). One FO used previously reported yields as a built-in edit check in a fruit CATI instrument. As very little specific fruit data are stored in Data Warehouse, the PRD were typically pulled from an FO dataset and imported into the necessary application. The most common use was for telephoning on paper. The PRD were printed either directly on the questionnaire or a special label. Several FOs utilized PRD for in-house CATI instruments with built-in edit checks. One FO reported pre-printing PRD on mailed questionnaires. All FOs said the various procedures worked well and seemed to improve the overall fruit data, but no objective tests had been conducted.

Potato Surveys: The third most common utilization of PRD in FOs was in potato data collection programs. The most common use was a pre-response prompt/verification of previously reported data, although two FOs reported using previous month's data in more of a balance sheet or post-response check for telephone data collection. Typical variables included previously reported acreage, yield/production, storage capacity and previous month's stocks. Several FOs utilized as much data as possible from the entire crop cycle in their final potato disposition survey. This included the previously noted production and capacity variables as well as monthly sales and stocks information. This was usually presented as a pre-response prompt, allowing respondents to verify or correct as needed the previous cycle's data. Typically data were pulled from various in-house data sets and populated into SMS for utilization into the respective application. Limited potato data, especially monthly variables, are stored in the Data Warehouse. Several FOs reported pre-printing data on mail questionnaires. PRD were also used in personal interviews and for telephone enumeration on paper. No FOs noted using a CATI instrument for potato data collection.

There are some unique caveats to note about FO potato programs: typically, these would be one or two page questionnaires, eliminating much of the printing barriers faced with other surveys. By the time the potato cycle wraps up, FOs have a fair amount of accumulated monthly data, much of which is asked again on the annual disposition. Potato producers tend to be larger, more intensely managed operations leading to frequent data requests, not only from NASS, but other entities as well. This lends credence to the idea of using PRD to reduce respondent burden,

increase utilization of data, verify existing data and collect any item missing data via future contacts. Most FOs commented that they thought utilizing PRD in their potato programs did reduce burden and improve respondent perceptions of subsequent requests.

Grain Prices/Off Farm Grain Stocks: While this was one of the more common uses of PRD reported by FOs, many of the usages seemed to be merely varying derivatives of those already provided via several applications by HQ.

Cattle on Feed Surveys: Several FOs reported using PRD in Cattle on Feed (COF) surveys. Note that this is in addition to the HQ developed uses of PRD already in the data collection process as a result of research conducted by Stanley and Safer (1997). The data were used as a pre-response prompt or for verification. Variables included capacity and previous month's inventory. This allowed FOs to use a balance sheet approach in data collection (previous month less outshipments plus placements) and verify that the results were consistent. Data were gathered from Data Warehouse or SAS datasets. Several methods were employed to differentiate between record level estimated data, although there was not any way to separate item level edited data. No FOs reported using PRD on mailed questionnaires; rather the PRD were pre-printed on paper questionnaires used in telephone or personal interviews. FOs indicated that COF PRD filled an important need in their monthly data collection.

County Estimates Surveys: Field offices also reported using PRD in their county estimates surveys. For all of these applications, Enhanced List Maintenance Operations (ELMO) data were incorporated into in-house CATI instruments as post-response checks. Several FOs incorporated built in edit checks, while others placed the PRD in the instrument and allowed enumerators flexibility in following up or probing on the differences. This allowed for simultaneous verification of large changes as well as ensuring new ELMO data were correct. FOs reported perceived benefits, but none had tested any of the effects.

3.2 FO Proposed Uses for PRD

FOs offered a variety of suggestions of possible surveys and variables which could utilize PRD. These possibilities ranged from fairly simple, specific items to broad, complex extremes. Following is a summary, in order of frequency mentioned by FOs:

“For what surveys and uses should NASS consider employing PRD?”

- Quarterly Agricultural Survey (QAS) follow-ons could utilize previously reported total land, crop land, storage capacity and crop acres in lieu of re-asking the question in follow-on surveys
- PRD could be used to create customized questionnaires for specific operations
- Ag Yield surveys could just verify previously reported yields and the planted acreage reported in the QAS
- Any variables that operations are asked multiple times over the course of a crop year
- All probability and non-probability surveys
- Previously reported planted acreages on any follow-on survey in the crop year
- Allow operators to verify PRD on the county estimates surveys

- Census follow-on surveys
- Any variable for which a current to current (C/C) estimator is calculated
- Provide previous year's PRD to large complex floriculture operations to allow for comparison/as an anchor for the current year
- Current total and crop acreages on the Agriculture Resource Management Survey
- Use November 1 Potato Survey information in lieu of re-asking it on the December QAS
- For vegetable surveys, especially in States with complex cropping systems and/or multiple crops grown in one year

Most of the FO comments on tools needed revolved around the gathering or consolidation of PRD and then importing it into the appropriate instrument. These suggestions covered a broad range of possibilities for the process. A few FOs also mentioned better tools to overcome more logistical type problems in incorporating PRD.

“What tools are needed to improve our ability to use PRD consistently and effectively?”

- Data from non-probability surveys needs to be stored in Data Warehouse (DW)
- Data into DW sooner following the survey
- Flags in DW for edited or estimated variables
- Ensure the transition or importation process from DW into SMS is standardized and works smoothly
- SMS needs the ability to pull data from other sources (rather than just DW)
 - Ability to access DW via the ELMO extract and import directly into SMS
- For FOs printing PRD on labels, an expanded selection of standardized label formats
- More automation across the entire process
- Total system to create customized questionnaires and import PRD
- Incorporation of Farm Service Administration (FSA) data for QAS follow-ons
- Enhanced abilities to print PRD in questionnaires, especially multi-page, pre-printed instruments (for enumerator reference and/or mail versions)
- More or easier access to PRD via current applications (e.g. enhanced “drill down” features) throughout the entire data collection process

Several FOs also acknowledged the wide range of considerations or precautions necessary in using PRD.

“Other concerns (about PRD)”

- If use of PRD is expanded, need to ensure standardized integration within surveys, across FOs
 - Existing difficulties integrating PRD into applications need to be overcome
- Need to prioritize ways to prevent inadvertently using edited, imputed or estimated data
- Confidentiality concerns:
 - Will giving back PRD foster a sense of “Big Brother” or raise concern amongst respondents about the amount of data collected and stored?
 - Preventing others (besides the respondent) from seeing PRD

- Not printing PRD in mail questionnaires, but rather using it as a post-response check in other modes
- Ensuring that the effort to reduce burden doesn't influence respondents or bias their responses

4. CONCLUSION AND RECOMMENDATIONS

NASS currently uses PRD in variety of national surveys, such as Agricultural Labor, Quarterly Agricultural Surveys and Agricultural Yield Surveys. PRD are primarily used for edit checks in HQ developed CATI instruments, but are also used to reduce contacts or in lieu of asking certain questions.

Additionally, FOs also reported using PRD for both pre-response prompts as well as post-response check/verification. Some FOs also provided PRD as background information or interview preparation for enumerators. Some FOs utilized PRD for larger, national surveys (such as QAS) but on a limited number of variables. 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 the data were used, not only for the current survey, but also as an important source of control data verification and updates (county estimates surveys). Most uses were for telephone data collection using paper instruments or CATI, or for field interviews. Few incidences of pre-printing on mail questionnaires were reported. PRD were typically imported from Data Warehouse or other datasets into SMS for dissemination into the respective application. There was usually no way to keep reported data separate from edited data. FOs generally reported positive experiences with PRD, but also noted difficulty in smoothly or easily incorporating the data.

FOs also expressed interest in employing PRD in a variety of situations. Suggestions ranged from one or two specific variables on certain surveys to completely customized questionnaires using all available PRD. Most suggested using PRD for pre-response prompts (verification) or to skip questions entirely in follow-on surveys. Field Offices also noted a need to improve or standardize procedures and applications for integrating PRD into data collection. Finally, 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.

In summary:

1. PRD have many potential applications which can reduce respondent burden while improving reporting consistency and respondent goodwill. However, there can be difficult hurdles to consistently incorporating PRD across multiple mode surveys. Caution must also be used to ensure confidentiality of the data and PRD also creates concern about the possibility of "satisficing" by respondents.
2. Based upon prior research, NASS utilizes PRD in a variety of data collection processes, primarily as a built-in edit check in CATI instruments. This allows enumerators to immediately confirm/resolve large differences or inconsistent reports. This usage is

consistent with the existing literature on PRD, and appears to be well received by the individual FOs at an operational level.

3. Many FOs reported using additional PRD in one form or another for a variety of surveys. Often these applications were structured as an attempt to reduce respondent burden. However, very few (if any) of the applications had been rigorously tested, some of the usages may not be consistent with NASS data collection parameters and there were some significant operational difficulties noted.

4.1 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 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:
 - 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.
 - 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 is 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?

That said, Web instruments, with their ability to collect paradata, or information about the response process, (Couper, 2005) may be a unique opportunity to collect and evaluate otherwise unavailable information on how respondents behave in reaction to PRD.

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. For instance, asking “How many acres of cropland do you have?” is not linguistically more concise than “According to our records, there are 2,000 acres of cropland in this operation. Is that correct?”

Skipping the planted question (for example, on an end of the year survey) and asking “According to our records, there were 1,000 acres of corn planted on this operation. Of these, how many were harvested for grain or seed?” creates several issues:

- It assumes that the respondent is the same and that the PRD are what they reported (as opposed to edited or incorrectly provided data).
 - It also presupposes that the 1,000 acres was correct or actually planted (Fowler, 1995) and may create an incentive to satisfice (Tourangeau et al, 2000): the respondent hears the question and thinks ‘well, I had intended to plant 1,000 acres of corn but it was too wet and 100 acres were shifted into soybeans’ and answers that 900 acres were harvested for grain while not clarifying the planted acres. This leads to additional complications when the soybean questions are asked and total acres are tabulated.
 - In conjunction with the presupposition issue, it borders on being a multi-faceted question, as it asks respondents to confirm planted acres and then provide harvested acres all in the same question (Fowler, 1995).
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). Research into mitigating seam effects by providing PRD asked about overlapping time periods and allowed respondents to change both previous and current period answers (Holmberg, 2004). Using this type of structure (e.g. on an Agricultural Yield survey) could ultimately lead to more inconsistency between survey periods.
 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. It would be similar to an SMS comment sheet, but with PRD on current survey variables. This would be an extension of the post-response check concept and provide office CATI enumerators or field enumerators and telephone enumerators calling with paper questionnaires respectively, with what had been reported for the same survey variables on previous surveys. Enumerators would need to be trained thoroughly in acceptable uses. Additionally, procedures must be in place to ensure that only final, reported data are utilized in the profile.

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. Operational concerns would be a minimum as these are one to two page surveys and FOs could pre-print PRD onto paper questionnaires for mailing and for telephone and personal interviewing. These commodities' estimates often rely on current/historical indications for acreage estimates. Respondents may be asked current and previous year's data on the instrument or current data is compared to previously reported data as part of the summarization process. Providing PRD back to respondents could increase consistency and comparability of an estimator already utilized. Additionally, these data are typically summarized differently than probability survey data. Any inadvertent measurement error resulting from utilizing PRD in a non-probability survey may have a minimal effect (compared to that in a probability survey) on final summarized data. FOs may already be utilizing a host of methods to collect data and derive estimates for specialty commodities. PRD appear to be an additional, useful tool in the process.
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. Appendix A provides an example of this type of analysis and develops some observations about the overall utility of expanded PRD use in data collection.

However, even if testing shows current data to be consistent with PRD, caution must be exercised in using PRD to skip follow-on questions. Consider the QAS cycle for a farmer in the Midwest:

Crop acreage and planted acreage must be asked in June. There are no PRD to utilize. (Even if he was in the March data collection as part of the previous year's sample, most Midwest farmers haven't begun any fieldwork in March. All of his intentions are just that—intentions. Total cropland acreage, let alone acreage of specific commodities, may not even be finalized for the year.) However, even during the June survey period, planting may not have been finished and there is still time to switch acreage between crops. Re-planting may be required after the data are collected. It's possible that September small grain planted acreage could be different from what was reported in June. Skipping planted acreage questions would miss this change. Harvested acres and yield or production still need to be asked. In December, planted acres must be asked or at least confirmed, in order to capture any possible changes from what was reported in June. During March, using the past year's PRD to skip planted acres questions (and even total and cropland questions) would ignore any expected changes for the new crop year. In the QAS cycle, there simply aren't that many questions in follow-on surveys that can be skipped as a result of utilizing PRD.

REFERENCES

- Biemer, Paul P. and Lyberg, Lars E. (2003). *Introduction to Survey Quality*, Hoboken, NJ: John Wiley and Sons.
- Couper, Mick P. (2005). "Technology Trends in Survey Data Collection." *Social Science Computer Review*, Vol. 23, No. 4, pp. 1-16.
- Dillman, Don A. and Smyth, Jolene D. (2007). "Design Effects in the Transition to Web-Based Surveys." *American Journal of Preventative Medicine*, May, pp. 90-96.
- Farrar, Martha and Beckler, Dan (2006). "Previously Reported Data (PRD) Action Plan." *Program Planning Council Meeting*, March.
- Fowler, Floyd J. (1995). *Improving Survey Questions: Design and Evaluation*, Thousand Oaks, CA: Sage Publications, Inc.
- Holmberg, Anders. (2004). "Pre-printing Effects in Official Statistics: An Experimental Study." *Journal of Official Statistics*, Vol. 20, No. 2, pp. 341-355.
- Kalton, Graham and Citro, Constance F. (1993). "Panel Surveys: Adding the Fourth Dimension." *Survey Methodology*, Vol. 19, No. 2, pp. 205-215.
- Kalton, G. and Schuman, H. (1982). "The Effect of the Question on Survey Responses: A Review." *The Journal of the Royal Statistical Society*, Vol. 145, No. 1, pp. 42-57.
- Krosnick, Jon A. (1991). "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology*, Vol. 5, pp. 213-236.
- Mathiowetz, Nancy A. and McGonagle, Katherine A. (2000). "An Assessment of the Current State of Dependent Interviewing in Household Surveys." *Journal of Official Statistics*, Vol. 16, No. 4, pp. 401-418.
- Stanley, Jaki S. and Safer, Martin A. (1997). "'Last Time You Had 78, How Many Do You Have Now?' The Effect of Providing Previous Reports on Current Reports of Cattle Inventories." *Proceedings of the American Statistical Association*, Survey Research Methods Section, pp. 875-880.
- Sudman, S., Bradburn, N.M., and Schwarz, N. (1996). "Answering a Survey Question: Cognitive and Communicative Processes." *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*, San Francisco: Josey-Bass Publishers.
- Tourangeau, Roger, Rips, Lance J. and Rasinski, Kenneth (2000). *The Psychology of Survey Response*, Cambridge, UK: Cambridge University Press.

APPENDIX A

Comparing PRD over the QAS Cycle

Chris Gottschall

1. INTRODUCTION

One of the more frequently suggested applications for utilizing previously reported data (PRD) is to ask, at the beginning of a crop year, what are generally thought to be relatively static items (e.g. total acres, cropland acres, grain storage capacity, etc.) and reuse these reported data over the current crop year as either a pre-response prompt or to skip re-asking the respective questions entirely over the crop year. The general consensus is that this would reduce respondent burden as well as possibly foster respondent goodwill. It could also possibly reduce spurious reports of change, making data more consistent over a crop year or survey cycle.

But would this really be the case? One of the primary concerns is that PRD would open up responses to satisficing: respondents simply agree with the PRD in order to minimize time expended on the survey, without regard or consideration of any possible changes to the data items. PRD must also be consistently utilized over the various modes the USDA's, National Agricultural Statistics Service (NASS) utilizes in order to avoid mode effects. Properly wording questions in order to utilize PRD as a pre-response prompt may result in wordier questions—reducing respondents' cognitive burden while increasing their time burden.

This current research focuses simply on assessing the validity of PRD as a substitute for re-asking certain questions over a crop year: how much do the “static” variables actually change, if at all? The results of these tests are intended to provide guidance into possible expanded uses of PRD or areas for further research. It is not the intent (nor is it possible, given the current data) to test or evaluate the social or cognitive issues previously mentioned. Rather, the goal is to use the available data to quantitatively evaluate the implications of using PRD in data collection.

2. METHODS

This analysis examined four quarters, an entire crop year, of Michigan Field Office Quarterly Agricultural Survey (QAS) final datasets (March 2007-December 2007). Sample sizes ranged from 1,318 in September 2007 to 1,851 in March 2007. The samples were drawn using a Multivariate Probability Proportion to Size (MPPS) design which is efficient and flexible. It also uses replicates and subpopulations to ensure comparability and consistency, while considering respondent burden. For the analysis, previously reported data were compared with future quarters' reports in an effort to measure the amount of change reported for several variables. Only records which had a response code (item code 9901) of “1-Complete” were included in the analysis. Additionally, only records which had a positive value for total acres operated (item code 900) for both quarters being compared were included in the respective iterations of the test, thus eliminating any known zeros, out of business records, etc. from the calculations (see Table 1).

The analysis looked at the unweighted data and examined the number of reports with different values and the mean amount of change for the variables across the survey periods. A possible extension of this research would be to utilize PRD in expanded population total indications and evaluate the differences; however, that was beyond the scope of the current analysis, which focused on measuring and analyzing the amount of change at the individual record level.

TABLE 1: Sample Sizes, Number Complete and Positive Total Acres

Survey Month	Sample Size	Number Complete (ic9901 = 1)	Positive Total Acres (ic900 not missing)
March 2007	1,851	1,341	1,286
June 2007	1,822	1,235	1,159
September 2007	1,318	1,045	1,003
December 2007	1,657	1,158	1,113

The current research utilized data that were previously reported in the QAS data collection of a preceding quarter in the current crop year. No control data (e.g. List Sampling Frame (LSF) data), other survey data (such as a current year livestock survey) or administrative data (FSA, etc.) were included. This was intended to provide as much direct comparability as possible and to keep the design of the analysis as simple and as straightforward as possible. An all-inclusive test utilizing as much PRD available from all sources could be designed and constructed in the future. This structure filled the need for a basic, preliminary analysis that would be a concise summary while still providing insight into the possible ramifications of using PRD.

3. RESULTS

Table 2 provides a summary of the data used for this research. Previously reported QAS data were compared to current QAS data for six time periods: March 2007 to June 2007; March 2007 to September 2007; March 2007 to December 2007; June 2007 to September 2007; June

TABLE 2: Records and Variables Analyzed

Quarters	Sample Size	Overlapped Records (total)	Overlapped Useable Records	Variables
1 March 2007 - June 2007	1,851 1,822	390	192	total acres, cropland acres, stor. cap., acres planted (corn, soybeans, wheat, oats)
2 March 2007 - September 2007	1,851 1,318	373	228	total acres, cropland acres, stor. cap., acres planted (wheat, oats)
3 March 2007 - December 2007	1,851 1,657	464	267	total acres, cropland acres, stor. cap., acres planted (corn, soybeans)
4 June 2007 - September 2007	1,822 1,318	808	496	total acres, cropland acres, stor. cap., acres planted (wheat, oats)
5 June 2007 - December 2007	1,822 1,657	878	478	total acres, cropland acres, stor. cap., acres planted (corn, soybeans) alfalfa hay
6 September 2007 - December 2007	1,318 1,657	1,272	817	total acres, cropland acres, stor. cap.

2007 to December 2007; and September 2007 to December 2007. Variables evaluated included: total acres, cropland acres, grain storage capacity, and planted acres for corn, soybeans, wheat, oats and alfalfa hay.

The following tables (Table 3 – Table 8) detail the results of the comparisons conducted on PRD for a given quarter to what was then reported in a subsequent quarter. Each table includes a list of variables, the mean and standard error (s.e.) for each quarter’s respective mean, the percent change between the two quarter’s means, the percent of records reporting a change between the two quarters and a summary of the magnitude of change between the time periods. The “Change Magnitude” column includes the quartile 1 (25 percent) and quartile 3 (75 percent) values to offer perspective on the amount of change between the reporting periods.

For all of the tables, the term “Change” can be considered interchangeable with “Difference” as there is no conclusive way to determine if the values are actually a change in the variable (e.g. the June total acres value actually increased or decreased from the March value); or, there was merely a difference between the reports (the respondent reported differing values for whatever reason—misspoke, incorrect answer, faulty cognitive heuristic, etc.). Regardless, the values reported in one quarter were not consistent with what was reported in another quarter—and measuring and evaluating these discrepancies was the overall goal of this analysis into PRD versus current reported data.

Table 3 compares what was reported in the March intentions survey to what the same operation reported in the June planted acreage survey.

Variable	Survey	Mean (se)	Percent Change in Mean	Records Reporting Change (%)	Change Magnitude (Q1, Q3)
Total Acres	March 2007	1,767 (162.4)	+ 2.7%	82.3%	-1, 20
	June 2007	1,815 (175.4)			
Cropland Acres	March 2007	1,608 (151.3)	+ 3.6%	77.1%	-28, 50
	June 2007	1,666 (167.9)			
Grain Storage Capacity	March 2007	112,805 (18,276)	- 0.8%	47.4%	-1000, 0
	June 2007	111,916 (18,592)			
Acres Planted:					
Corn	March 2007	783.5 (95)	+ 6.1%	68.8%	-10, 28
	June 2007	831.6 (110.6)			
Soybeans	March 2007	829.6 (128.9)	- 1.3%	42.2%	-53, 20
	June 2007	819.2 (127.4)			
Wheat	March 2007	238 (28.0)	- 6.7%	31.3%	-6.5, 0
	June 2007	222 (26.2)			
Oats	March 2007	55 (8.0)	- 7.3%	26.0%	-14, 5
	June 2007	51 (7.1)			

There were 192 overlapped and useable records (operation in both samples, both surveys coded as “complete” and total acres present for both reports) used in the comparison. While there are some differences in the means, none of them are significant. However, some of the percentages of change are fairly large, most notably corn and wheat. The change in corn mean acres is

possibly due to acreage plans not being finalized and/or acres being shifted into corn for weather related reasons later in the planting season. However, wheat acres are not that easily explained: both the March and June questions ask about winter wheat acres seeded in the fall of the preceding year (2006). Possibly, by June, respondents had replanted winter killed wheat acres, leading to the change in mean acres.

Table 4 examines the data from the March intentions survey and the September small grains survey. There were 228 overlapped and useable records used in the comparison. Interestingly, wheat planted acres differ much less than in the preceding March to June comparison. The difference in grain storage capacity could be explained by possible construction over the summer months.

TABLE 4: March PRD vs. September Reported Data

Variable	Survey	Mean (se)	Percent Change in Mean	Records Reporting Change (%)	Change Magnitude (Q1, Q3)
Total Acres	March 2007	1,445 (131.9)	+ 2.7%	75.4%	-40, 31
	September 2007	1,484 (143.5)			
Cropland Acres	March 2007	1,355 (129.7)	+ 2.7%	78.9%	-25, 40
	September 2007	1,392 (141.9)			
Grain Storage Capacity	March 2007	104,684 (16,178)	+ 4.4%	46.5%	-1000, 800
	September 2007	109,330 (16,093)			
Acres Planted:					
Wheat	March 2007	239 (29.3)	+ 1.7%	27.6%	-1.5, 5
	September 2007	243 (27.9)			
Oats	March 2007	52 (7.2)	+ 1.9%	26.8%	-13, 2
	September 2007	53 (9.5)			

Table 5 compares the March intentions survey to the December year-end survey. There were 267 overlapped and useable records used in the comparison. The difference in corn acres is somewhat smaller than that noted in the March versus June survey (Table 3), while soybean planted acres are consistent with the March to June comparison.

TABLE 5: March PRD vs. December Reported Data

Variable	Survey	Mean (se)	Percent Change in Mean	Records Reporting Change (%)	Change Magnitude (Q1, Q3)
Total Acres	March 2007	1,648 (128.9)	+ 1.7%	77.2%	-30, 50
	December 2007	1,677 (135.3)			
Cropland Acres	March 2007	1,531 (126.2)	+ 3.4%	77.2%	-29, 61
	December 2007	1,583 (133.6)			
Grain Storage Capacity	March 2007	111,057 (16,004)	+ 2.2%	47.2%	-250, 1000
	December 2007	113,475 (14,639)			
Acres Planted:					
Corn	March 2007	776 (79.5)	+ 4.6%	68.2%	-28, 48
	December 2007	812 (88.9)			
Soybeans	March 2007	762 (105.5)	- 2.1%	35.9%	-53, 17
	December 2007	746 (104.7)			

Table 6 compares the June planted acreage to the September small grains survey. There were 496 overlapped and useable records used in the comparison. The oat planted acreage differs by over 13 percent, especially interesting as the vast majority of oats are planted by June 1. Wheat acres (again, planted the previous fall) also increased for comparable records. Total and cropland acre changes in direction and magnitude are consistent with what was observed in the preceding comparisons.

TABLE 6: June PRD vs. September Reported Data

Variable	Survey	Mean (se)	Percent Change in Mean	Records Reporting Change (%)	Change Magnitude (Q1, Q3)
Total Acres	June 2007	1,067 (78.5)	+ 3.5%	64.7%	-11, 20
	September 2007	1,104 (82.4)			
Cropland Acres	June 2007	986 (77.3)	+ 1.5%	64.5%	-11, 20
	September 2007	1,001 (77.9)			
Grain Storage Capacity	June 2007	65,476 (8,621)	- 1.1%	36.1%	0, 0
	September 2007	64,750 (8,235)			
Acres Planted:					
Wheat	June 2007	179 (15.5)	+ 7.8%	26.8%	-1, 2
	September 2007	193 (18.4)			
Oats	June 2007	45 (7.7)	+ 13.3%	16.1%	-3, 0
	September 2007	51 (8.9)			

Table 7 examines the values reported in June and December. There were 478 overlapped and useable records used in the comparison. For the first time in the analysis, total and cropland acres declined from the PRD. The question structure for alfalfa hay is slightly different in December from June. This may explain some of the difference in reported data.

TABLE 7: June PRD vs. December Reported Data

Variable	Survey	Mean (se)	Percent Change in Mean	Records Reporting Change (%)	Change Magnitude (Q1, Q3)
Total Acres	June 2007	1,133 (84.1)	- 1.4%	77.2%	-30, 20
	December 2007	1,117 (79.1)			
Cropland Acres	June 2007	1,047 (82.6)	- 1.3%	81.4%	-30, 25
	December 2007	1,033 (77.0)			
Grain Storage Capacity	June 2007	73,674 (9,819)	+ 1.7%	40.8%	0, 1000
	December 2007	74,910 (9,160)			
Acres Planted:					
Corn	June 2007	548 (55.7)	+ 1.1%	56.5%	-5, 7
	December 2007	554 (54.9)			
Soybeans	June 2007	491 (58.9)	- 2.9%	39.1%	-10, 10
	December 2007	477 (57.5)			
Alfalfa	June 2007	116 (9.8)	-11.2%	43.5%	-19, 15
	December 2007	103 (9.8)			

And finally, Table 8 compares September PRD with December reported data. Of the comparisons made across time periods, there are the least number of variables present in both the September and December surveys. However, there were 817 overlapped and useable records

used in the comparison. Similar to the June and December comparison, the current data for both total and cropland acres are smaller than the PRD. Grain storage capacity also shows the second largest percentage increase for any of the comparison periods.

TABLE 8: September PRD vs. December Reported Data

Variable	Survey	Mean (se)	Percent Change in Mean	Records Reporting Change (%)	Change Magnitude (Q1, Q3)
Total Acres	September 2007	902 (55.6)	- 2.1%	72.1%	-18, 18
	December 2007	883 (51.7)			
Cropland Acres	September 2007	834 (54.5)	- 0.8%	77.1%	-16, 25
	December 2007	827 (51.4)			
Grain Storage Capacity	September 2007	59,580 (6,177)	+ 4.2%	35.3%	-3000, 275
	December 2007	62,089 (6,377)			

Table 9 presents a summary of the frequencies of comparisons, the range of percent changes in the means, the range of percent of record reporting changes and the range of the magnitudes of change for the variables of interest tested in the analysis.

TABLE 9: Summary of Results

Variable	Number of Comparisons	Range of:		
		Percent Change in Means	Records Reporting Change (%)	Change Magnitude (Q1, Q3)
Total Acres	6	-2.1%, +3.5%	64.7%, 82.3%	-40, 50
Cropland Acres	6	-1.3%, +3.7%	64.5%, 81.4%	-30, 60
Grain Storage Capacity	6	-1.1%, +4.4%	35.3%, 47.4%	-3000, 1000
Acres Planted:				
Corn	3	+1.0%, +6.1%	56.5%, 68.8%	-28, 48
Soybeans	3	-2.9%, -1.2%	35.9%, 42.2%	-53, 20
Wheat	3	-6.7%, +7.8%	26.8%, 31.3%	-6.5, 5
Oats	3	-5.6%, +13.3%	16.1%, 26.8%	-14, 5
Alfalfa	1	-11.2%	43.5%	-19,15

From this, there does appear to be a fair amount of difference across time periods for these variables when PRD are compared to current period survey data. However, none of the differences proved significant. The percent change in the unweighted means, while not overly large (except for a few commodities and time periods), are roughly similar to the expected coefficient of variation for the final weighted estimate. This leads back to the underlying questions surrounding PRD: 1) How much additional error, if any, would using PRD create for a given survey and over an entire crop year? 2) Would this be a negligible, ignorable amount or would it adversely affect the precision and quality of the estimates?

After reviewing the results of the Michigan dataset analysis, the findings on overlap, completion rates, etc. were extrapolated forward to the US level. This was an additional effort to further quantify the utility of expanded PRD usage, and was intended to serve as an approximation of the maximum amount of PRD available for use in a subsequent quarter:

March PRD for use in June:

75.3 percent complete – 3.4 percent Other and Estimated – 4.4 percent Known Zero = 67.5 percent

67.5 percent - 1.2 percent Mail Edited (7.9 percent mail completes * 15 percent (approx.) hand-edited) = 66.3 percent

66.3 percent - 6.6 percent Reported Data not equal to Data Warehouse (66.3 percent * 10 percent (approx.) edited during the interactive edit)

59.7 percent of the March Records for which PRD are potentially available

59.7 percent PRD potential * 25 percent overlapped with June = 14.9 percent of June records which could use March PRD.

Of these records, approximately 12-15 questions are asked in both March and June (variables for which March PRD could be utilized in some manner for June). The actual number realized could vary somewhat by state. This includes six questions about total operation acreage (owned, rented, etc.), one question about grain storage capacity and five to eight questions about acres planted for various crops. There are approximately 60 total questions asked in June.

June PRD for use in December:

70.8 percent complete – 2.4 percent Other and Estimated = 68.4 percent

68.4 percent - 0.9 percent Mail Edited (6.4 percent mail completes * 15 percent (approx.) hand-edited) = 67.5 percent

67.5 percent - 6.8 percent Reported Data not equal to Data Warehouse (67.5 percent * 10 percent (approx.) edited during the interactive edit)

60.7 percent of the June Records for which PRD are potentially available

60.7 percent PRD potential * 50 percent overlapped with December = 30.4 percent of December records which could use June PRD.

Of these records, approximately 18-20 questions are asked in both June and December (variables for which June PRD could be utilized in some manner for December). The actual number realized could vary somewhat by state. This includes 6 questions about total operation acreage (owned, rented, etc.), 1 question about grain storage capacity and 11-13 questions about acres planted for various crops. There are approximately 75 total questions asked in December.

4. CONCLUSION AND RECOMMENDATIONS

This analysis wasn't intended as an exhaustive, detailed analysis, but rather to provide some insight into the actual behavior and comparability of PRD vs. current reported data. It was designed to be a simple summary analysis which would provide a synopsis of the amount of change in PRD vs. current survey period reported data over a complete crop year.

There were some important limitations in this analysis:

- There were a fairly small number of useable records (overlapped and non-missing) between respective time periods. The number of useable records ranged from 192 records to 817 per period, or just under 11 percent of the sample to 49 percent of the sample.
 - Similarly, variables with missing values presented an additional confounding factor: how should these be interpreted when they were present and positive in the PRD? Does missing equal none or were they actually don't know, refusals, etc.
- There were only a relatively small number of variables present across all the surveys that could be tested.
- This experiment was not set up as a test—it involved analyzing already collected data. The lack of a split sample or control group in this case probably doesn't warrant concern.
- This test only used data from other QAS. It could have possibly used other current crop year survey or Data Warehouse data. However, there may not be much additional crop data (other than QAS data) available in time to use as PRD during a crop year.

However, the results of this preliminary test found (for a given crop year on a large, national survey):

- A limited number of overlapped records.
- A minimal number of variables were present from quarter to quarter to utilize PRD.
- A fair amount of change or difference in the reports.

That said, if we do pursue an expanded usage of PRD in data collection, further research needs to be conducted into its impact on data quality. Currently, there are simply too many unknowns. In order to provide defensible procedures and estimates, these possible implications must be thoroughly assessed in advance of any extensive changes in the current PRD usage paradigm.

Specifically, further research should:

1. Use PRD to re-summarize the survey data in an attempt to measure the impact of PRD on survey indications. However, this creates several fairly significant difficulties (missing data, etc.) and it seems unlikely to completely isolate possible PRD effects.
2. Repeat this test using all of the national QAS data for an entire crop year. This would be a much larger group of records and hopefully would result in a larger percentage of total reports to compare over the different time periods.
3. Conduct a split sample test. This would require utilization of PRD for a random portion of a given QAS sample while the other records in the sample are collected via current standard procedures. The two groups are then compared to isolate any effect of PRD on data quality. This procedure could face limitations similar to the first recommendation.

However, from these preliminary findings, it appears that the current NASS utilizations of PRD (predominately as CATI edit checks) may actually be the most efficient use of PRD on large, national probability surveys within the agency.

APPENDIX B

Memo as sent to FOs on February 15, 2008 requesting information on FO level usage of PRD

TO: State Directors
Deputy Directors
All NASS Field Offices

THROUGH: Marshall Dantzler
Deputy Administrator For Field Operations

Robert T. Bass
Director
Census and Survey Division

FROM: Martha Farrar
Chief
Data Collection Branch

SUBJECT: Field Office Uses of Previously Reported Data in Data Collection

**ACTION: Submit Requested Information in the Attached Spreadsheet to HQ CSD
DCB mailbox**

DUE DATE: COB, March 14, 2008

The Data Collection Standards Working Group was chartered in 2007 with Martha Farrar as Business Council Liaison and Bob Bass as Executive Sponsor. Group members are Edwin Anderson, Patrick Boyle, Emily Caron, Sammye Crawford, Morgan Earp, Pat Evans, Chris Gottschall and Gail Gregory. The working group was entrusted with the responsibility of ongoing review of virtually all aspects of NASS data collection. The purpose of the group is to ensure that data collection instruments, processes, methodology, etc., are sound, effective and non-biasing, and that respondent burdened is lessened to the greatest degree reasonably possible.

At their initial meeting, the group was asked make reviewing current uses of Previously Reported Data (PRD) in data collection a top priority. The goal behind this charge was to identify and evaluate current PRD usage and to expand the use of successful, methodologically sound practices. This charge to the group was issued prior to the release of PSM-47-07 which dictates that proposed PRD usage be evaluated to determines that it meets our respondent burden and data quality criteria.

In 2006, a list of the most widespread HQ uses of PRD in data collection was compiled and presented to the PPC. The PPC members were pleased with the extent of PRD used in CATI and recommended DCB incorporate the similar uses of PRD in Electronic Data Reporting (EDR) as appropriate. In 2005, DCB and SAB asked FOs to report how they were utilizing PRD, including the surveys in which PRD were being used and the methodology behind such use. The

Data Collection Standards Working Group reviewed these data and identified areas and surveys to evaluate for research and implementation.

To ensure that all possible candidates are considered, the group is now asking each FO to report their current (or most recent) uses of PRD in the data collection process as well as any uses they would like to make of PRD in the future. The attached spreadsheet has tabs for current (or recent) uses and for proposed uses.

For current (recent) uses, states should include detailed information on the usage of PRD: the survey(s), the specific questions (or types of questions), how the data are used, the training provided to enumerators in using the data with respondents, whether the PRD were reported data or (possibly) edited data, the effect on respondent burden, any specific PRD data handling procedures, and how effective the FO believes the practice is. All of the PRD usage questions may not apply, but please be as thorough as possible in describing procedures. Examples are included in the spreadsheet.

For proposed uses of PRD, please also mention any tools needed to improve the use of PRD in Data Collection. A response from every FO is needed by March 14, regardless of whether the FO is currently using (or has recently used) PRD in data collection. If you have questions, please call Martha Farrar (202-720-2837).

For all surveys, FOs are encouraged to provide Survey Administration Branch (SAB) with as much input as possible (via Evaluation forms) on data collection issues and questionnaire issues: wording, routing, respondent problems, enumerator problems, etc. The Data Collection Standards Working Group will work with Survey Administrators to review the E-2s. State input is vital in identifying problem areas. Especially where questionnaires are concerned, identifying problems, pre-testing alternatives and recommending solutions are key ways the Data Collection Standards Working Group can help minimize respondent burden.