United States Department of Agriculture

National Agricultural Statistics Service

Research and
Development Division
Washington DC 20250

RDD Research Report
Number RDD-09-01

September 2009

Using Previously Reported Cropland Acreage in Data Collection

HoaiNam N. Tran
Michael W. Gerling

EXECUTIVE SUMMARY

The National Agricultural Statistics Service (NASS) conducts surveys on the United States' and Puerto Rico's agriculture for the purpose of estimating crops, livestock, production practices, farm economics, and related items of interest.

In an effort to reduce respondent burden, NASS' Research and Development Division examined the relationship between the June and September Agricultural Survey 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. Both the five-state and all-state results for cropland are documented in this report. A subsequent report will encompass more of the items for which the use of previously reported data has potential.

The total cropland analyses show 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 improvement. One of the ways this could be accomplished is by including additional critical edit checks on cropland data during the editing process.

RECOMMENDATIONS

1. In the context of an overall review of the potential increased use of previously reported data with the agency's surveys, additional items (beyond the cropland focused on in this report) should be studied to determine whether data quality could be adequately preserved by carrying forward June Agricultural Survey responses to the September Agricultural Survey. The continued research should focus on land owned, land rented to, land rented from, and storage capacity.

2. If additional items are so identified for which carrying forward data would not jeopardize data quality, consider reducing respondent burden for the September Agricultural Survey by carrying theses items forward and not re-asking the questions in September. Multiple questionnaire items would need to be carried forward for the resulting respondent burden reduction to be sufficient to make this use of previously reported data practical and cost effective. The research documented in this report indicated that this would only by possible, even for cropland, in 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).

    a. If future research identifies additional questionnaire items that can be carried forward from the June Agricultural Survey into the September Agricultural Survey, the following are also recommended:

        i. 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.

        ii. Do not carry forward cropland acreages from June Agricultural Survey into the September Agricultural Survey for the following twenty-one states: Alabama, Arizona, California, Colorado, Florida, Kansas, Kentucky, Maryland, Missouri, Nevada, New Jersey, New Mexico, New York, North Carolina, Oklahoma, Oregon, South Carolina, Tennessee, Texas, West Virginia, and Wyoming.

3. Add a critical edit check to the Survey Processing System (SPS) Edit to flag records reporting relatively large differences in cropland compared to historical data.

4. Office survey administrators should reemphasize the importance of collecting accurate data at the field and office enumerator training workshops and to the office staff responsible for reviewing and editing the questionnaires.

Using Previously Reported Cropland Acreage in Data Collection

HoaiNam N. Tran
Michael W. Gerling[/1]

Abstract

The National Agricultural Statistics Service (NASS) conducts surveys on the United States' and Puerto Rico's agriculture for the purpose of estimating crops, livestock, production practices, and farm economics.

NASS' Research and Development Division examined the agency's June and September Agricultural Surveys to determine whether identical questions asked on both surveys could be eliminated on the September survey for June respondents, thereby reducing respondent burden. The questionnaires for these two surveys for five states (Delaware, Iowa, Nebraska, Texas, and Washington) were reviewed and identical questions (cropland, land owned, land rented to, land rented from, and storage capacity) were identified. This report specifically focuses on cropland. A second report focusing on additional items is scheduled for late 2009.

Initially, between-quarter differences in cropland for the same operation were measured. If the indicated variation in the answers was not statistically significant, it was concluded that the questions could be omitted in September for those operators responding in June. This initial test showed that four of the five test states could have the June cropland data carried forward to the September Agricultural Survey.

The analyses were then expanded to include all states participating in the June and September Agricultural Surveys. The results indicated that if the quality of the cropland data collected in June is improved then using the June reported data in September is feasible in 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.

KEY WORDS: Previously Reported Data, Agricultural Survey, Data Collection

[/1] HoaiNam N. Tran and Michael W. Gerling are Mathematical Statisticians with the National Agricultural Statistics Service - Research & Development Division, located at 3251 Old Lee Highway, Fairfax, VA 22030.

1. INTRODUCTION

The National Agricultural Statistics Service's (NASS) primary mission is to provide timely, accurate, and useful statistics on the United States' and Puerto Rico's agriculture. NASS conducts hundreds of agricultural surveys annually.

The Quarterly Agricultural Survey Program is conducted four times each year, in March, June, September, and December, with June being the base quarter. Each quarter, survey participants are asked to report acres of cropland operated as of a stated reference date. In the past, respondents have questioned NASS' rational in asking some of the same questions each quarter. Hence, a research project was constructed to examine the practicality of omitting repetitive questions and thereby reducing respondent burden.

The goals of this research project were to:

1. Determine if the cropland question could be omitted in September for operations who were sampled for both quarters and responded to the question in June. If so, the June responses could be carried forward to the September data set for processing and analyses. This procedure would reduce respondent burden to some extent since the length of the September instrument would be reduced for those operations responding in June.

2. Examine the impact of carrying June cropland data forward to the September data set as described in goal number 1.

3. Document ways to improve the data collection and analysis processes as discovered by this research.

2. JUNE AND SEPTEMBER AGRICULTURAL SURVEYS

The Quarterly Agricultural Survey Program obtains agricultural data from farmers and ranchers across the nation to estimate crop production, grain storage supply, and livestock and poultry production. All agricultural operations across the United States are eligible to be selected in the sample. The annual sample sizes for the United States for years 2003 through 2005 ranged from 50,000 to 90,000 operations. See Appendix A for sample sizes by state and year. The survey questions typically refer to either the first day of the survey month or to a specific crop year as reference points.

Data collection is a multi-mode combination of mail, telephone interview, Computer Assisted Telephone Interview (CATI), personal enumeration, and Web data collection.

2.1 CROPLAND

For this report, cropland acreage was the item examined. On both the June and September Agricultural Survey questionnaires, the precise wording of the question on cropland is "*Of the total acres operated, how many acres are considered cropland, including land in hay, summer fallow, cropland idle, cropland used for pasture and cropland in government programs?*" The reference date is June 1st or September 1st, depending on the survey period.

This question is one of a series of questions concerning the operator's land. Questions pertaining to land owned, rented to others and rented from others are asked to assist the operator in deriving the total acres in the operation.

Total cropland is used in imputing missing individual crop acreage values for inaccessible and refusal records. It is also utilized in the calibration process of the Quarterly Agricultural Survey data. NASS, however, does not publish a cropland estimate based on the June or September data.

In general, an agricultural operation's total cropland should not significantly change between June and September. In June, most crops have been planted or are being planted and any land rental arrangements have been finalized. By September, crops are either growing or being harvested. While total cropland can change between June and September if an operation purchases or sells some acreage during this time period, such valid mid-season acreage changes are rare.

3. ANALYSIS PROCESS EXPLAINED

Initially, five states (Delaware, Iowa, Nebraska, Texas, and Washington) were selected to be analyzed. Delaware was chosen due to its relatively small number of operations, which make the data sets manageable for testing purposes. The other states were selected based on their use of standardized questionnaires and their collective representativeness of U.S. agriculture.

Questions occurring on both the June and September Agricultural Survey questionnaires were identified.

All list operations sampled in June and September were analyzed with the following restrictions:

1. Complete, usable reports were obtained for both quarters.

2. Operations which were inaccessible, refused to complete the survey, or provided estimated data were removed from the study.

3. Operations which responded as "out-of-business" were also eliminated.

Data plots (Figure 1, page 7) were created from the total cropland information for the remaining operations to obtain a visual picture of the distribution of the data to identify and examine any

outliers. A review of these plots revealed extreme outliers that required additional study. Those outlier data points determined to be in error were removed from the study.

F-tests, linear regression analyses, 95% confidence intervals, and percent differences of expanded data were used to determine whether or not the reported June and September cropland acreages were statistically different.

3.1 TESTING THE DIFFERENCES BETWEEN JUNE AND SEPTEMBER RESPONSES USING LINEAR REGRESSION

The regression model under consideration was $y = ax + b$, where x and y are the June and September values, respectively. For this study, June and September data always have a non-trivial linear relationship, meaning that the slope, *a*, is never zero (and, in fact, should be close to 1) at the designated confidence level. A 95 percent confidence level was used in this study. The hypotheses for the slope *a* and intercept *b* are as follows:

$$H_o : a - 1 = 0 \qquad \text{and,} \qquad H_o : b = 0$$
$$H_a : a - 1 \neq 0 \qquad\qquad H_a : b \neq 0$$

If a sampled operation's responses are not significantly different between June and September, the estimated intercept has to be small (near zero) and the estimated slope near 1. This causes the formula ($y = ax + b$) to become approximately $y = 1x + 0$ (or $y = x$).

3.2 TESTING THE DIFFERENCES BETWEEN JUNE AND SEPTEMBER RESPONSES USING AN F-TEST

For testing the differences between the quarterly responses, let J_{ijk}, S_{ijk}, and W_{ijk} be the June data, September data, and September weight respectively for operation i, state *j*, and year *k*, where i is indexed over the number of matching records. Further, letting $y_{ijk} = S_{ijk} - J_{ijk}$, the F-test p-value of the regression model $y_{ijk} = J_{ijk}$ would also tell the relationship between the June and September data. If the resulting p-value is smaller than α, then the matching data items for June and September are significantly different. P-values closer to 1 indicate less dissimilarity between the June and September values.

3.3 EXAMINING THE AGGREGATE EFFECT OF CARRYING JUNE CROPLAND DATA FORWARD

To measure the effect of carrying June data forward to September, historical data sets were used

and the process was simulated for 2003, 2004, and 2005. The corresponding September weights for each year were used on this simulated data set to obtain new expanded totals. The expanded total from the simulated September records was then compared to the original expanded September total. This procedure was first applied for the five test states then applied to all the surveyed states (43 states) for 2003, 2004, and 2005. Alaska and Hawaii were excluded since they do not participate in either the June or September Agricultural Survey. Connecticut, Massachusetts, New Hampshire, Rhode Island, and Vermont were also excluded since they do not participate in the September Agricultural Survey.

Let J_{ijk}, S_{ijk}, and W_{ijk} be defined as above. The two expanded totals were compared by considering the percent change in state j and year k with the following statistic: $p_{jk} = 100 * \frac{\sum_{ijk} S_{ijk} W_{ijk} - \sum_{ijk} J_{ijk} W_{ijk}}{\sum_{ijk} S_{ijk} W_{ijk}}$.

The comparison measured the potential impact of re-using the data from the good June reports for September, rather than re-asking the questions.

3.3.1 ABSOLUTE UPPER CONFIDENCE INTERVAL LIMIT

Definition: The absolute upper confidence interval limit of the expanded percent differences, $(p_{1k},..., p_{jk},..., p_{nk})$ for the year k, is defined as follows:

$$\bar{p}_k + t_{\alpha/2} \sqrt{\frac{\sum_{j=1}^{n} (|\, p_{jk} \,| - \bar{p}_k)^2}{n(n-1)}}, \text{ where } \bar{p}_k = \frac{\sum_{j=1}^{n} |\, p_{jk} \,|}{n}.$$

Taking the absolute values of the expanded differences allows an assessment of the overall magnitude of the aggregate June to September differences, without considering the direction of any individual changes that might have occurred. Hence, this procedure reveals any changes that otherwise would be masked by similar but opposite expanded differences elsewhere.

The resulting absolute upper confidence interval limit was compared to the absolute value of the expanded percent difference for each state/year combination. The two possible conclusions are:

a. If the absolute percent difference is less than its respective absolute upper confidence interval limit, then there is not a significant difference between the June and September values for that state and year.

b. If the absolute percent difference is greater than or equal to the respective upper confidence interval limit, then the difference may not be accounted for by randomness. Therefore, a closer review of the June and September cropland data would need to be conducted to see why the cropland acreages were different.

3.3.2 T-VALUES

The t-values used throughout the study are based on the sample sizes (See Appendix A.) from the respective states, and are shown in Table 1 below. Delaware has different t-values since it has small and varying sample sizes. In comparison, Iowa, Nebraska, Texas, and Washington sample sizes are larger than 1,500, so the t-value of 1.96 was sufficient. The t-values for these four states differ only at the thousandth decimal place, which was considered negligible for this study.

Table 1: T-Values Used For Each State

State(s)	Year	Degrees of Freedom (df)	$t_{df,\alpha/2}$
Delaware	2005 2004 2003	96 104 63	1.99 1.98 2.00
Iowa Nebraska Texas Washington	All Years	>300	1.96

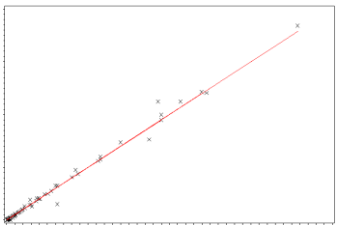
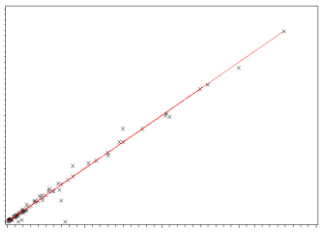
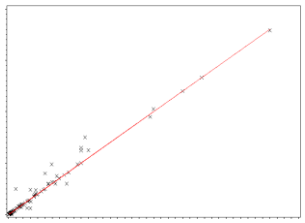
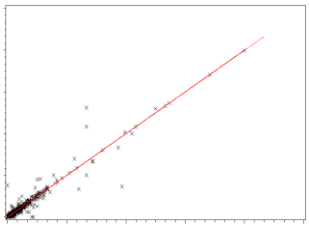
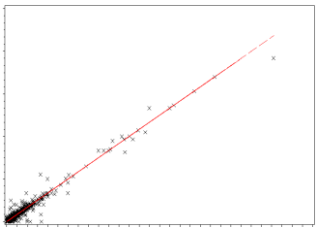
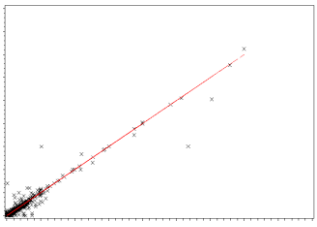
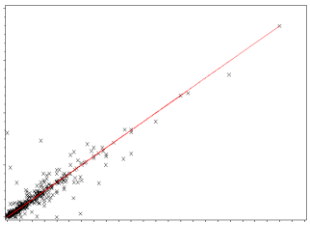
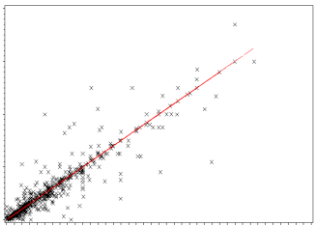
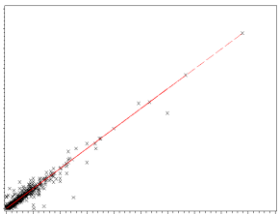
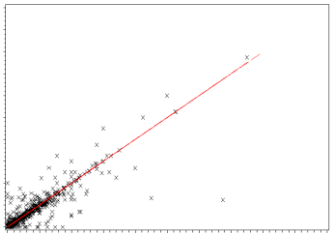
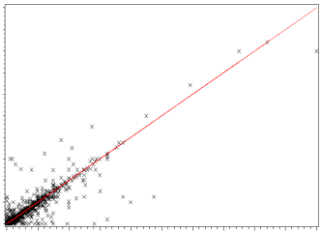
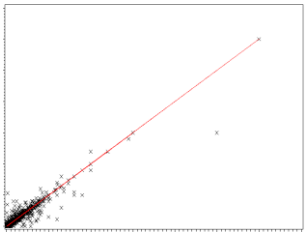
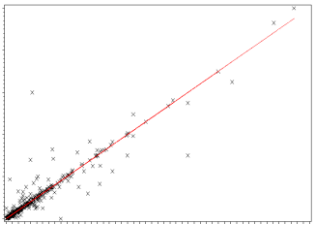
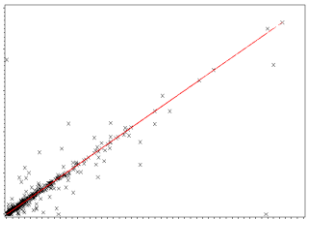
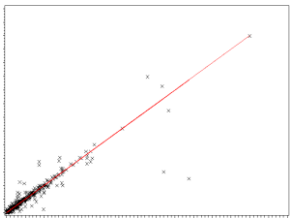
4. CROPLAND ANALYSES

Figure 1, located on the next page, displays data plots for the years of 2003, 2004, and 2005 for the five test states. The scales of the horizontal and vertical axes are the same for each plot. The red line represents the function y = x, where x and y are June and September total cropland, respectively.

If the operation's values were the same from June to September then all of the points, denoted in black, would fall on this line. If a point was far away from the line then its values were reviewed for legitimacy and points representing data errors were removed.

Visually, all of the data points trend around the line $y(x) = x$. However, the 2004 Nebraska data and the 2004 and 2005 Texas data had more points lying further away from the line than the other state/year combinations.

Figure 1: June Cropland Acres versus September Cropland Acres by Operation [/1]

State & Year	2003	2004	2005
DE			
IA			
NE			
TX			
WA			

[/1] The horizontal (June data) and vertical (September data) axes are the same for each plot.

Table 2, on the following page, displays the F-test, slope, intercept, confidence intervals for slope and intercept, expanded percent difference, and the absolute expanded percent difference upper confidence interval (C.I.) limit for the five test states in 2003, 2004, and 2005.

Overall, the data sets that passed all the tests were Delaware's 2003 and 2004 data, all three years of data for Iowa, and the 2003 Texas data. The p-values for these six data sets ranged from 0.07 to 0.74 and were greater than the alpha level. The slope estimates were also very close to 1, ranging from 0.97 to 1.03. The intercept estimates ranged from -14.66 to 34.90. While statistically some of these estimates are significantly different from zero (the ideal intercept value for these regressions), they are all small relative to the scale of the data and are practically insignificant. Both the slope and intercept confidence intervals contain the values of 1 and 0, respectively.

The absolute expanded percent difference between the two quarters' estimates for these passing state/year combinations ranged from 0.03 to 2.52, which were statistically insignificant compared to their respective absolute upper confidence interval limits. Thus, this implies that the June Agricultural Survey cropland data for Delaware in 2003 and 2004; for Iowa in all three years; and for Texas in 2003 could have been brought forward to the corresponding September's quarter without dramatically affecting data quality.

Before moving forward on the instances that passed these initial statistical tests, a closer examination of the occurrences that failed them was conducted. An examination of the data plots for these failed occurrences (Figure 1) shows that a majority of the records are located on or near the line $y = x$. Also, records that are not located near the line of symmetry are generally located above the line, meaning that these respondents reported more cropland in September than in June. However, a review of the historical data showed some of these increases were due to errors in the data.

Table 2: F-Test, Slope, Intercept, Slope and Intercept Confidence (C.I.) Intervals, Expanded Percent Difference, and Absolute Expanded Percent Difference Upper Confidence Interval Limit

State/ Year		F-test P-value	Slope	Intercept	Slope 95% C.I.	Intercept 95% C.I.	Expanded Percent Difference	Absolute Expanded % Difference Upper C.I. Limit
DE	2003	0.06	1.03	-14.46	(-62.02, 33.10)	(0.99, 1.05)	0.71	3.05
	2004	0.45	0.99	-11.15	(-54.30, 32.01)	(0.97, 1.02)	2.52	4.38
	2005	0.04	1.02	29.62	(-16.80, 76.04)	(0.99, 1.05)	-10.50	6.51
IA	2003	0.07	1.03	-14.66	(-50.02, 20.70)	(0.99, 1.05)	2.24	3.05
	2004	0.25	1.01	-0.11	(-21.13, 20.92)	(0.99, 1.02)	-1.15	4.38
	2005	0.74	0.99	18.84	(-30.71, 68.39)	(0.96, 1.02)	0.03	6.51
NE	2003	<0.0001	0.93	65.10	(23.01, 107.10)	(0.91, 0.96)	-4.03	3.05
	2004	0.0002	0.95	34.25	(-2.17, 70.65)	(0.93, 0.97)	1.06	4.38
	2005	<0.0001	0.96	40.20	(12.09, 68.32)	(0.94, 0.97)	1.20	6.51
TX	2003	0.14	0.97	34.90	(-26.06, 95.87)	(0.94, 1.01)	1.35	3.05
	2004	<0.0001	0.83	168.78	(123.27, 214.30)	(0.80, 0.85)	2.68	4.38
	2005	<0.0001	0.89	114.03	(79.16, 148.90)	(0.88, 0.91)	-0.02	6.51
WA	2003	0.01	0.96	99.95	(20.10, 179.79)	(0.93, 0.98)	-0.91	3.05
	2004	<0.0001	0.87	195.57	(103.60, 287.54)	(0.83, 0.90)	5.88	4.38
	2005	<0.0001	0.90	141.32	(66.10, 216.53)	(0.88, 0.93)	0.95	6.51

Table 3 displays some of the outliers with erroneous data. The reasons for the data being in error are also provided in the table.

Table 3: Outliers with Erroneous Data

State	Year	Record	June Cropland	September Cropland	Reason in Error
DE	2005	1	800	3,800	Based on historical data, another operation's acreage may have been incorrectly included in Sept.
NE	2003	2	0	4,855	Based on historical data, June cropland acreage should have been around 4,800 acres.
NE	2003	3	125	2,850	Based on historical data, operation's cropland in Wyoming is sometimes included.
TX	2004	4	24,500	11,000	Based on historical data, 2004's June's cropland data are incorrect.
TX	2004	5	90,000	5,700	One operator overseeing two operations. Data for first operation were mistakenly collected and posted to second operation and vice versa.
TX	2004	6	2,500	20,000	Based on historical data, June cropland acreage should be around 20,500 acres.
WA	2004	7	0	15,000	Based on historical data, there are multiple operations. June cropland acreage should be around 15,000 acres.
WA	2004	8	2,000	20,200	Based on historical data, September cropland acreage should be around 2,020 acres.
WA	2005	9	1,715	12,000	Based on historical data, there are multiple operations sometimes being included in the total cropland.
WA	2005	10	1,600	5,600	Based on historical data, September cropland acreage should be around 1,600 acres.

Table 4 summarizes the number of operations with outlier cropland data found to be in error. The erroneous data were removed from the test data set and the analyses were re-run.

Table 4: Number of Operations Whose Cropland Data Were Identified as Outliers, Found to be Incorrect, and Excluded from the Analyses

State	Number of Outliers Excluded		
	2003	2004	2005
DE	0	0	1
IA	0	0	0
NE	5	6	6
TX	0	10	13
WA	3	5	3

Table 5 displays the revised results of the five test states. The footnote (1/) indicates those particular states' data sets where no erroneous data were detected. In these cases the results are identical to those presented earlier in Table 2.

Table 5: F-Test, Slope, Intercept, Slope and Intercept Confidence Intervals, Expanded Percent Difference, and Absolute Expanded Percent Difference Upper C.I. Limit Without the Erroneous Outlier Data

State/Year	F-test's P-value	Slope	Intercept	Slope 95% C.I.	Intercept 95% C.I.	Expanded Percent Difference	Absolute Expanded % Difference Upper C.I. Limit
DE 2003[/1]	0.06	1.03	-14.46	(-62.02, 33.10)	(0.99, 1.05)	0.71	3.05
DE 2004[/1]	0.45	0.99	-11.15	(-54.30, 32.01)	(0.97, 1.02)	2.52	4.38
DE 2005	0.14	1.01	19.00	(-23.75, 61.75)	(0.99, 1.05)	-1.50	1.52
IA 2003[/1]	0.07	1.03	-14.66	(-50.02, 20.70)	(0.99, 1.05)	2.24	3.05
IA 2004[/1]	0.25	1.01	-0.11	(-21.13, 20.92)	(0.99, 1.02)	-1.15	4.38
IA 2005[/1]	0.74	0.99	18.84	(-30.71, 68.39)	(0.96, 1.02)	0.03	6.51
NE 2003	0.002	0.95	41.23	(8.90, 73.57)	(0.93, 0.97)	1.40	1.85
NE 2004	0.19	0.98	17.34	(-15.32, 49.99)	(0.96, 1.01)	0.44	1.84
NE 2005	0.12	0.99	20.61	(-1.53, 42.74)	(0.98, 1.00)	-0.15	1.52
TX 2003[/1]	0.14	0.97	34.90	(-26.06, 95.87)	(0.94, 1.01)	1.35	3.05
TX 2004	<0.0001	0.95	56.14	(18.88, 93.40)	(0.93, 0.97)	0.50	1.84
TX 2005	<0.0001	0.96	53.41	(24.22, 82.61)	(0.95, 0.98)	-1.88	1.52
WA 2003	0.13	0.98	44.64	(-10.78, 100.07)	(0.96, 1.01)	-0.46	1.85
WA 2004	0.14	0.98	63.37	(-7.65, 134.39)	(0.95, 1.01)	-0.88	1.84
WA 2005	0.18	0.99	28.66	(-13.20, 70.53)	(0.97, 1.01)	-0.23	1.52

[/1] Denotes the results were unchanged from the Table 2. No erroneous outlier data were removed from the data set.

By removing the erroneous outlier data, Delaware, Iowa, and Washington data passed the tests for all three years. Their F-test p-values were all larger than 0.05. Also, their slope and intercept confidence intervals contained both 1 and 0. The maximum absolute estimated percent difference between the June and September acreage data was 2.52 percent, corresponding to the percentage change of 2004 Delaware data. These differences were small, compared to their respective absolute upper confidence interval limits. These factors imply that the June cropland data could have been carried forward into September for those three states.

For Nebraska, the 2004 test results were improved substantially. The F-test p-value increased from 0.0002 to 0.19. Both 1 and 0 were contained in the slope and intercept confidence intervals, respectively. Also, the percent difference between the two quarters dropped to 0.44 percent.

Excluding the erroneous outlier data did not significantly improve Nebraska's 2003 results. There were too many operations whose cropland acreage differed by +/-3,000 acres between June and September, causing the data to fail the F-test. However, the resulting estimated slope (0.95), intercept (41.23), and the absolute percent difference (1.40) were marginally close enough to cautiously justify using the June cropland data in September.

For Texas, there were several records with significant fluctuations in cropland acres between June and September. Without going back to the operations, it was impossible to conclude in several cases, whether or not the data were erroneous. The significant differences in the 2003 and 2004 data caused the F-test to fail. Also, the 2004 absolute percent difference between the two quarters (1.88) was larger than its respective absolute upper C.I. limit (1.50). Further research is needed to determine whether Texas' June cropland data can be carried forward into September.

Overall, the tests for the five selected states only implied that the data could be carried forward for certain states.

4.2 CROPLAND ANALYSIS EXPANDED

Next, the expanded percent differences in cropland between June and September Agricultural Surveys for all 43 states that participated in both surveys were considered. Reviewing Tables 6 and 7 (pages 14 and 15) along with the results of the initial five test states, there were 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) whose expanded percent differences between June and September data were less than the respective absolute upper C.I. limit. This indicates that the June data could be carried forward into the September Agricultural Survey for these states.

On the other hand, there were twenty-one states (Alabama, Arizona, California, Colorado, Florida, Kansas, Kentucky, Maryland, Missouri, Nevada, New Jersey, New Mexico, New York, North Carolina, Oklahoma, Oregon, South Carolina, Tennessee, Texas, West Virginia, and

Wyoming) whose expanded percent differences between June and September data were greater than or equal to the absolute upper C.I. limit in at least one year. This indicates that bringing June data forward would have an adverse effect on the data. Therefore, carrying June cropland data forward into September is not recommended for these states.

Table 6: Expanded Cropland Percent Difference Between June and September Data by State and Year

State/Year	Expanded Percent Difference Between June and September Data		
	2003	2004	2005
Alabama	7.91/1	1.79	-2.55
Arizona	17.95/1	1.45	1.42
Arkansas	-1.48	1.87	-0.19
California	-0.48	7.87/1	4.06/1
Colorado	11.53/1	-0.17	-1.10
Delaware	0.71	2.52	-10.50/1
Florida	1.08	1.55	8.99/1
Georgia	-1.86	-1.80	0.01
Idaho	0.21	0.61	-1.90
Illinois	3.25	-0.43	0.98
Indiana	-0.27	0.18	1.31
Iowa	2.24	-1.15	0.03
Kansas	4.82/1	0.14	-1.55
Kentucky	4.43	4.07/1	-1.53
Louisiana	1.83	0.75	0.90
Maine	0.36	-0.03	0.57
Maryland	-2.53	3.09/1	1.44
Michigan	-0.87	1.88	2.12
Minnesota	-0.74	1.01	-1.09
Mississippi	0.93	0.75	1.41
Missouri	4.94/1	5.43/1	-1.20
Montana	-0.89	-1.51	0.14
Nebraska	-4.04	1.06	1.20
Nevada	-3.74	-3.70/1	-1.49
New Jersey	-2.48	-4.17/1	1.01
New Mexico	3.27	-1.54	-6.25/1
New York	0.38	-0.77	3.71/1
North Carolina	0.64	4.40/1	0.13
North Dakota	1.03	0.83	0.86
Ohio	1.76	2.49	0.83
Oklahoma	5.27/1	1.83	4.69/1
Oregon	5.88/1	1.04	2.08
Pennsylvania	3.98	0.56	-1.54
South Carolina	3.88	-0.03	-6.78/1
South Dakota	2.01	0.21	3.06
Tennessee	2.90	1.98	3.40/1
Texas	1.35	2.68/1	-0.02
Utah	2.57	-0.37	-0.14
Virginia	0.97	-1.21	1.69
Washington	-0.91	5.88/1	0.95
West Virginia	6.01/1	0.13	-3.98/1
Wisconsin	0.38	1.76	1.33
Wyoming	-12.48/1	3.15/1	-2.01

/1 Absolute percentage change is greater than or equal to the absolute upper C.I. limit

Table 7: The Averages, Standard Errors, and Absolute Upper C.I. Limits of Cropland Expanded Percent Difference between June and September Data, by Year

Type of Statistics	Year		
	2003	2004	2005
Average	1.67	1.07	0.11
Absolute Average	4.73	2.54	3.14
Weighted Average	1.52	1.20	0.50
Absolute Standard Error	4.53	2.32	3.16
Absolute Upper Confidence Interval Limit	4.61	2.62	3.08

5. IMPROVING DATA QUALITY

The data collection and data review processes need to be improved prior to carrying the June cropland data forward to matching records in the September Agricultural Survey. Some sources for potential error in the cropland data include misunderstanding by the respondent as to what is included or excluded as cropland, data entry mistakes, illegible handwriting on the questionnaires, and confusion or misunderstanding between the respondent and the field enumerator. All of these need to be addressed in order to minimize error in the cropland data.

Furthermore, a critical edit check that would flag those records in the June Survey whose cropland data differed substantially from historical data should be added to the June Agricultural SPS Edit.

Also, a table in NASS' Interactive Data Analysis System (IDAS) showing all operations with positive cropland values in June that was also sampled for September needs to be developed. The table would also display the operations' expansion weights for September, their reported cropland for June, and their contributions to the overall indications by multiplying these two items. NASS statisticians would then be able to see those operations having the most impact on September's cropland if the June cropland values are carried forward.

Finally, field office survey administrators need to reemphasize the importance of collecting accurate data at the field and office enumerator training workshops and to the office staff responsible for reviewing and editing the questionnaires.

6. CONCLUSION

Once the quality of the cropland data is improved, it would be feasible to carry forward June cropland data to the September Agricultural Survey for 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. Multiple questionnaire items would need to be carried forward for the resulting

respondent burden reduction to be sufficient to make this use of previously reported data practical and cost effective. If ultimately implemented, periodic data quality assessment should be conducted periodically to measure the quality of any June data brought forward and to validate this use of previously reported data.

For the other states participating in the June and September Agricultural Surveys, future data quality checks need to be conducted to see if June cropland data become stable enough to be used in September.

This study shows that using previously reported data for items that "*seem*" to be fairly stable can adversely affect the quality of the estimates. Only through careful study and research can the impact of using previously reported data be measured.

The Research and Development Division will continue to examine using previously reported data as a way to reduce respondent burden while improving the quality of data.

7. REFERENCES

Myers, Raymond and Walpole, Ronald 1972, "*Probability and Statistics for Engineers and Scientists*", The Macmillan Company.

Lohr, Sharon 1999, "*Sampling: Design and Analysis*", Brooks/Cole Publishing Company.

Appendix A

June Agricultural Survey and September Agricultural Survey

Sample Sizes and Matching Records

For Years

2003, 2004 and 2005

By State

June Agricultural Survey and September Agricultural Survey Sample Sizes and Matching Records

St	June Ag. Sample			September Ag. Sample			Matching					
							2003		2004		2005	
	2003	2004	2005	2003	2004	2005	No.	%	No.	%	No.	%
AL	998	1,288	1,268	662	971	950	384	58%	596	61%	572	60%
AZ	419	462	426	389	476	456	249	64%	315	66%	295	65%
AR	2,132	2,092	2,022	1,477	1,535	1,439	970	66%	1,001	65%	942	65%
CA	1,939	2,133	2,103	1,247	1,642	1,540	868	70%	1,153	70%	1,095	71%
CO	1,600	1,905	1,973	943	1,188	1,175	686	73%	866	73%	864	74%
DE	308	372	341	212	281	273	159	75%	206	73%	203	74%
FL	462	567	552	433	599	591	241	56%	356	59%	360	61%
GA	1,305	1,568	1,562	1,050	1,270	1,243	599	57%	779	61%	770	62%
ID	1,645	1,848	1,771	1,144	1,455	1,346	840	73%	1,102	76%	1,013	75%
IL	2,289	2,455	2,389	2,169	2,290	2,171	1,442	66%	1,522	66%	1,434	66%
IN	1,861	2,191	2,195	1,884	2,094	2,063	1,203	64%	1,363	65%	1,376	67%
IA	1,908	2,258	2,388	1,867	2,297	2,471	1,060	57%	1,367	60%	1,508	61%
KS	4,006	4,257	4,226	2,383	2,754	2,719	1,578	66%	1,842	67%	1,810	67%
KY	1,549	1,784	1,783	1,318	1,295	1,310	816	62%	795	61%	814	62%
LA	1,276	1,488	1,389	824	1,021	985	527	64%	673	66%	660	67%
ME	281	302	311	256	291	303	222	87%	256	88%	261	86%
MD	695	1,001	932	451	733	711	337	75%	535	73%	504	71%
MI	1,607	1,908	1,885	1,155	1,497	1,431	757	66%	982	65%	951	66%
MN	2,172	2,400	2,359	1,835	2,125	2,046	1,181	64%	1,404	66%	1,370	67%
MS	1,221	1,536	1,499	932	1,350	1,283	612	66%	940	70%	903	70%
MO	2,303	2,718	2,707	2,083	2,486	2,465	1,346	65%	1,627	65%	1,626	66%
MT	1,500	1,777	1,828	1,510	1,959	1,926	1,024	68%	1,344	69%	1,404	73%
NE	2,580	2,944	3,015	1,851	2,362	2,333	1,261	68%	1,630	69%	1,665	71%
NV	222	258	273	185	207	218	119	64%	135	65%	142	65%
NJ	305	428	453	213	409	416	145	68%	275	67%	283	68%
NM	569	778	762	368	660	601	242	66%	464	70%	455	76%
NY	1,083	1,257	1,221	689	887	850	445	65%	566	64%	542	64%
NC	1,553	1,609	1,590	1,149	1,425	1,286	801	70%	961	67%	918	71%
ND	2,249	2,371	2,366	1,609	1,788	1,833	1,197	74%	1,327	74%	1,381	75%
OH	1,493	1,683	1,668	1,232	1,442	1,443	811	66%	962	67%	958	66%
OK	2,031	2,378	2,365	1,577	2,072	2,120	1,042	66%	1,385	67%	1,420	67%
OR	764	1,006	1,006	496	765	745	357	72%	542	71%	557	75%
PA	1,297	1,467	1,436	983	1,092	1,085	664	68%	767	70%	752	69%
SC	806	1,214	1,179	669	1,143	1,098	435	65%	756	66%	726	66%
SD	2,002	2,527	2,490	1,770	2,219	2,130	1,202	68%	1,557	70%	1,516	71%
TN	1,624	1,724	1,790	1,068	1,146	1,140	684	64%	750	65%	729	64%
TX	3,871	4,511	4,257	2,913	3,770	3,419	1,910	66%	2,570	68%	2,376	69%
UT	495	779	747	449	903	812	320	71%	610	68%	565	70%
VA	1,131	1,505	1,440	703	1,116	1,052	438	62%	745	67%	706	67%
WA	1,510	1,764	1,798	1,218	1,514	1,571	874	72%	1,070	71%	1,157	74%
WV	512	559	524	398	534	506	231	58%	315	59%	292	58%
WI	1,886	2,116	2,105	1,790	1,986	2,010	1,101	62%	1,262	64%	1,286	64%
WY	764	1,013	988	468	729	650	298	64%	479	66%	429	66%