Assessing an Administrative Data Source as a Sampling Frame

Wendy J. Barboza[1], Bill Iwig[2]
[1]USDA, NASS, 3251 Old Lee Highway, Room 305, Fairfax, VA 22030
[2]USDA, NASS, 1400 Independence Ave, SW, Washington, DC 20250

Abstract
The National Agricultural Statistics Service (NASS) conducts hundreds of surveys every year and prepares reports covering virtually every aspect of U.S. agriculture. To accomplish this, NASS maintains a list frame of producers (and agribusinesses), which includes previously reported crop data. For many years, NASS has partnered with USDA's Farm Service Agency (FSA) to use their data as an administrative data source because most producers report their planted crop acreages to FSA on an annual basis. Since FSA data are more current than data on NASS' list frame, NASS launched an operational pilot program in the state of Nebraska to investigate using FSA administrative data as a sampling frame. Although results from the pilot program did not meet survey objectives, various operational programs within NASS continue to utilize FSA administrative data, just not as a sampling frame.

Key words: sampling frame, administrative data, NASS, FSA

1. Introduction

The U.S. Department of Agriculture's (USDA) National Agricultural Statistics Service (NASS) conducts hundreds of surveys every year and prepares reports covering virtually every aspect of U.S. agriculture. NASS maintains a list frame containing names, addresses, telephone numbers, and other descriptive data on producers (and agribusinesses) and an area frame covering all land area in the U.S. To form multiple frame survey indications[1], data collected from the list frame sample are combined with data collected from the area frame operators who are not on the list frame. In this respect, the area frame accounts for the incompleteness of the list frame. This methodology ensures that every producer has a chance of selection.

For many years, NASS has partnered with USDA's Farm Service Agency (FSA) to use their data as an administrative data source since most producers report their planted crop acreages to FSA on an annual basis. Prior to 1997, NASS researched using FSA data as a sampling frame and administrative data source. After passage of the Federal Agriculture Improvement and Reform Act of 1996 (the 1996 Farm Bill), which reduced the emphasis on producers reporting to FSA, NASS temporarily stopped working on this issue. When the Farm Security and Rural Investment Act of 2002 (the 2002 Farm Bill) required producers participating in the FSA program to report all crops, NASS once again pursued examining different options for using FSA administrative data.

The Quarterly Agricultural Survey (QAS) is conducted once every three months in March, June, September, and December. Data collected by the QAS are used to set national- and state-level commodity estimates for planted acres, harvested acres, production, and on-farm grain stocks. In December 2004, the QAS survey design was utilized to conduct a research study in two states, Nebraska and Minnesota, to evaluate the possibility of using FSA administrative data as a sampling frame (instead of the NASS list sampling frame). Two additional states, North Carolina and Oregon, were added to the research study in June 2005. The survey indications and coefficients of

[1] NASS refers to macro-level data from a survey as an indication, rather than an estimate. This distinction is used because published numbers are typically based on multiple sources. For example, more than one survey may be conducted or administrative data may be used in conjunction with the survey results in producing official published estimates.

variation produced from the research studies were compared to those calculated from the operational program. The evaluation showed mixed results, as the performance of the survey indications and coefficients of variation varied by state and commodity.

Since there was sufficient motivation for using FSA administrative data as a sampling frame in terms of respondent burden reduction, operational efficiency, and quality of survey indications, NASS subsequently made a decision to initiate an operational pilot program in one state. NASS' target population is small compared to those of many other government agencies, resulting in respondent burden being more of a concern with NASS surveys than elsewhere. With the limited population size, NASS often surveys the same operators multiple times during its annual survey cycle. Consequently, the agency wanted to reduce respondent burden for producers who had already reported their planted crop acreages to FSA or who were being contacted multiple times by various organizations. In addition, NASS hoped this new methodology would reduce data collection costs in the long term. Finally, it was speculated that the reliability of the survey indications would significantly improve since the FSA data are more current than data on NASS' list frame.

In November 2005, an intra-agency team was established to identify all of the surveys impacted by this decision and to prepare the specifications for integrating the methodology into these surveys. Beginning in December 2006, NASS started using FSA administrative data as a sampling frame in the state of Nebraska. After two years, the operational pilot program was discontinued because the objectives were not sufficiently achieved.

This paper provides an overview of NASS assessing an administrative data source as a sampling frame. Although there were a number of surveys affected by the new methodology, this paper will only focus on the QAS. The QAS methodology for both sampling frames will be covered along with the modifications implemented after the first year of the pilot program. In addition, the results from an analysis comparing preprinted administrative data to reported survey data will be presented as well as a discussion on the reliability of the survey indications. In conclusion, the reasons for discontinuing the operational pilot program will be explained.

2. QAS Methodology Using the NASS List and Area Frames

Using the NASS frames, the survey cycle for the QAS starts in June (i.e., the base month) and ends in March of the following year. Multiple frame survey indications are produced by combining data from the list and area frame samples.

As stated earlier, NASS maintains a list frame containing names, addresses, telephone numbers, and other descriptive data on producers. For the list frame, the sampling unit is a producer on NASS' list and the reporting unit is any farm associated with the producer. So, when a sampled producer is involved in multiple operating arrangements, separate questionnaires are completed for each one. In early spring, the list frame population is "frozen" to select the sample. Strata are defined based on the control data (i.e., previously reported data) for the producers. These strata are different for each state, and are not design strata. Instead, they are only used to define the population, identify prob-1 (i.e., certainty) sampling units, and adjust for nonresponse. In 2005, strata in Nebraska were formed using total cropland, calculated cropland, on-farm grain storage capacity, and hay acres.

Reduced-list sampling methodology is used to define the list population; this means that producers with small amounts of cropland or capacity are excluded from the sampled population and are represented by the area frame component. Producers in the "large" strata are identified as prob-1 selections and are included in the sample every quarter. These are not eligible for nonresponse adjustment and the data for any refusals and inaccessibles among these selected units must be manually imputed. In order to target a multitude of commodities, Multivariate Probability Proportionate to Size (MPPS) sampling (Kott and Bailey, 2000) is used to sample the remaining reduced-list population and the approach involves sampling producers from the following components: general, row crops, and small grains. During sample selection, the control data are used to target the commodities of interest for each of the three components. As with the strata formation, the targeted commodities are different for each state. In 2005, the targeted

commodities in Nebraska were calculated cropland, on-farm grain storage capacity, and reported cropland for the general sample; alfalfa hay, corn, dry beans, chickpeas, all hay, calculated row crop acres, other hay, sunflowers, soybeans, proso millet, and sorghum for the row crops sample; barley, oats, rye, calculated small grain acres, and winter wheat for the small grains sample. Once the samples for the three components are identified, the selected units are assigned to one of three replicates. The component and replicate number determine the quarters for which the sampled producer is interviewed. This rotation scheme is shown in Table 1.

Table 1: QAS Rotation Scheme

Sample Component	Replicate Number			
	June	September	December	March
General	1	1, 2, 3	1, 2, 3	2, 3
Small Grains	1, 2	1, 2, 3	1, 2, 3	2, 3
Row Crops	1	1	1, 2, 3	2, 3

After the list sample is selected, the sampled producers within each quarter are calibrated to state-level commodity totals based on the control data of the reduced-list population. This calibration is performed to adjust for procedures related to the implementation of MPPS sampling. After this procedure, each sampled producer is assigned an adjusted sampling weight. Note that a producer is in sample for multiple quarters, but their sampling weight is different for each quarter.

Data collected from the list frame sample are combined with data collected from the area frame operators who are not on the list frame. In June, NASS conducts an agricultural survey of geographic segments (i.e., areas of land approximately 1 square mile in size) sampled from the area frame. For the area frame, the sampling unit is a segment and the reporting unit is an area of land inside the segment that is operated under one type of arrangement. The reporting unit, referred to as a tract, may consist of agricultural land, non-agricultural land with agricultural potential, or non-agricultural land. Farm-level data are collected for the agricultural tracts. These tracts are matched against the reduced-list frame population based on name/address information and the non-matches comprise the Not-On-List (NOL) component. In June, the sampling weight for NOL agricultural tracts is adjusted by a tract-to-farm weight and the farm-level data are combined with those from the reporting list frame units to produce multiple frame survey indications. For other quarters, both the NOL agricultural tracts and non-agricultural tracts with agricultural potential are post-stratified based on data collected in June. A sample is selected and the original sampling weight is adjusted accordingly. Farm-level data are collected every remaining quarter (i.e., September, December, and March) from the associated operators using the same questionnaire as used for the list frame sample.

3. QAS Methodology Using the FSA Administrative Data

Using the FSA administrative data, the survey cycle for the QAS started in December (i.e., the base month) and ended in September of the following year. The reason for changing the base month from June to December was because most producers do not report their planted crop acreages to FSA until later in the calendar year. Similar to using the NASS frames, multiple frame survey indications were produced by combining data from the list and area frame samples.

The FSA list population was constructed using three files: the FSA Geographic Information System (GIS) data layer, the FSA 578 administrative data file, and the FSA name/address file. The FSA GIS data layer contained the geographic boundaries for all FSA farms; this file was also used to calculate the total acres of the FSA farm. The FSA 578 administrative data file included the current-year planted crop acreages by FSA farm number for producers who reported to FSA. The FSA name/address file consisted of name and address information for owners, operators, and others associated with the FSA farm. In August, NASS received the FSA files and the list population was considered "frozen" to select the sample. The first two files were matched using state, county of administration, and FSA farm number. The FSA GIS data layer was considered the primary file, so farms in the FSA 578 administrative data file which did not match to the FSA GIS data layer file were discarded from the list population. Farms on the FSA GIS data layer file which did not match to the FSA 578 administrative data file contained data

for total acres only. For the list frame, the sampling unit was an FSA farm number and the reporting unit was the parcel of land identified by the geographic boundaries. So, if the sampled FSA farm number changed, the questionnaire was completed based on the geographic boundary associated with the original FSA farm number. Similar to using the NASS list frame, strata were defined based on the administrative data for the FSA farms and were used to define the population, identify prob-1 sampling units, and adjust for nonresponse. In 2005, strata in Nebraska were formed using total land (instead of total cropland), calculated cropland, and hay acres. On-farm grain storage capacity was not used because it is not reported to FSA. Usually, this target variable is used to obtain data for crops stored on the operation (i.e., stocks).

Reduced-list sampling methodology was not used to define the FSA list population (i.e., all FSA farms were included). FSA farms in the "largest" stratum were identified as prob-1 selections and were included in the sample every quarter. These were not eligible for nonresponse adjustment and the data for any refusals and inaccessibles among these selected units were manually imputed. The remaining list population was sampled using MPPS sampling and the approach involved sampling producers from the following components: general, row crops, and small grains. During sample selection, the control data were used to target the commodities of interest for each of the three components. The targeted commodities in Nebraska were the same as those used for the NASS list frame with the exception of using total land instead of total cropland and the unavailability of on-farm grain storage capacity. Once the samples for the three components were identified, the selected units were assigned to one of three replicates. The component and replicate number determined which quarters the sampled unit was to be interviewed. The rotation scheme was unchanged from the traditional approach shown earlier. After the list sample was selected, the sampled units within each quarter were calibrated to state-level commodity totals based on the control data of the list population and assigned an adjusted sampling weight. The FSA farms in the list sample were then matched to the FSA name/address file. Since multiple names could be associated with the FSA farm, the primary contact was determined by ranking the names based on whether the person was listed as an owner, operator, other, or combination thereof.

Weighted data from FSA farms sampled from the list frame were combined with the NOL component, which was determined by overlaying the FSA GIS data layer with the NASS GIS data layer for the geographic segments sampled from the area frame in June. Any areas within the segments that were not covered by the FSA GIS data layer were identified, "digitized", and compared to the tract-level data collected in June. The NOL component consisted of all digitized areas within a segment which were classified as either an agricultural tract or a non-agricultural tract with agricultural potential. For the NOL component, the sampling unit and the reporting unit were the same and were referred to as a parcel of land, whether it was an entire tract or a "partial" tract (i.e., areas of land within the tract). These NOL parcels of land were assigned the sampling weight from the June Agricultural Survey (since all were selected for the sample) and parcel-level data were collected every quarter from associated operators using the same questionnaire as used for the list frame sample.

Since the reporting unit for both the list and area frame components was a parcel of land, questionnaires included a map on the front page and were mailed to all sampled units. This visual aid assisted the respondent in recognizing the FSA farm number that was selected from the list frame component or the NOL parcel of land identified from the area frame component.

4. Comparison of FSA Administrative Data to Reported Survey Data

In addition to making methodological changes to the QAS, NASS implemented numerous modifications to operational survey procedures. One major difference was preprinting the planted crop acreages (when available) for list units on the December questionnaire and instructing the respondents to change incorrect values. December was the only month where this procedure was performed since the planted crop acreages reported by the producers should not have changed between August and December. Planted chickpea acres were not preprinted because the questionnaire collected information by small and large, but the FSA administrative data were not available by these categories.

A detailed analysis was performed in December 2006 on the QAS respondents. Although chickpeas were not preprinted, the reported survey data for the small and large categories were summed together and included in the analysis. Since planted crop acreages were preprinted for all sampled list units rather than a subset of them (i.e., an experimental design was not conducted), the results from the analysis were used for informational purposes only and no statistical comparisons were made.

The evaluation showed the difference between FSA administrative and reported survey planted crop acreages varied by commodity. Focusing on the macro-level data, the unweighted values for both data sources were aggregated for responding list units. A percentage difference was calculated by dividing the difference between the FSA and reported totals by the FSA total. Negative percentages indicated the reported total was greater than FSA and positive percentages indicated the reported total was less than FSA. Table 2 shows the percentage differences by commodity.

Table 2: Comparison of Macro-Level Data in December 2006

Commodity	% Difference between FSA and Reported Data
Planted Acres:	
Irrigated Corn	-0.83%
Non-Irrigated Corn	-0.96%
Irrigated Sorghum	-1.97%
Non-Irrigated Sorghum	1.02%
Irrigated Soybeans	0.01%
Non-Irrigated Soybeans	0.10%
Chickpeas	45.27%
Dry Beans	0.56%
Proso Millet	-5.40%
Oil Sunflowers	-2.59%
Non-Oil Sunflowers	-2.14%
Irrigated Alfalfa Hay	-3.24%
Non-Irrigated Alfalfa Hay	6.75%

The absolute percentage difference was the lowest for soybeans, both irrigated and non-irrigated, and the highest for chickpeas, although chickpea totals were based on fewer than 10 reports. Besides chickpeas, the absolute percentage difference was greater than 5 percent for proso millet and non-irrigated alfalfa hay. It is worth mentioning that the percentage difference was less than 1 percent for irrigated and non-irrigated corn, a major commodity, as well as dry beans, which is considered a minor commodity. For sorghum and alfalfa hay, the percentage differences for irrigated acres were negative, while those for non-irrigated acres were positive. This could be due to producers changing non-irrigated to irrigated acres for those commodities.

Focusing on the micro-level data, the unweighted planted crop acreages for both data sources were compared for responding list units. This analysis excluded those for which both the FSA and reported data were zero or missing. Similar to the macro-data analysis, a percentage difference was calculated by dividing the difference between the FSA value and the reported survey value by the FSA value, except that 100 was added to each data source so minor differences did not result in large percentage differences when the FSA value was small. Again, negative percentages indicated the reported value was greater than the FSA value and positive percentages indicated the reported value was less than the FSA value. Table 3 displays the percentage differences by commodity grouped into categories as well as the number of reports included in the calculations.

Table 3: Comparison of Micro-Level Data in December 2006

Commodity	# of List Reports*	% Difference between FSA and Reported Data				
		< -50%	< 0% to -50%	No Change	> 0% to 50%	> 50%
Planted Acres:						
Irrigated Corn	781	2.2%	2.8%	89.8%	5.0%	0.3%
Non-Irrigated Corn	847	1.9%	3.9%	90.1%	4.0%	0.1%
Irrigated Sorghum	109	6.4%	5.5%	74.3%	11.0%	2.8%
Non-Irrigated Sorghum	492	1.4%	4.9%	84.6%	8.5%	0.6%
Irrigated Soybeans	392	1.5%	2.6%	92.3%	2.3%	1.3%
Non-Irrigated Soybeans	623	1.0%	1.4%	92.1%	5.3%	0.2%
Chickpeas	8	0.0%	37.5%	25.0%	37.5%	0.0%
Dry Beans	234	0.9%	3.4%	89.7%	5.6%	0.4%
Proso Millet	176	8.0%	8.5%	75.6%	5.7%	2.3%
Oil Sunflowers	105	2.9%	2.9%	82.9%	10.5%	1.0%
Non-Oil Sunflowers	50	4.0%	2.0%	82.0%	12.0%	0.0%
Irrigated Alfalfa	348	4.0%	7.2%	79.9%	8.3%	0.6%
Non-Irrigated Alfalfa	630	2.9%	7.6%	75.6%	11.9%	2.1%

* Excludes reports for which both the FSA and reported data were zero or missing.

For the column labelled 'no change', the large (small) percentage differences corresponded to the small (large) percentage differences at the macro-level. The percentage difference in this column was the highest for soybeans, both irrigated and non-irrigated, and the lowest for chickpeas. Mimicking the previous analysis, the percentage difference in this column was small for proso millet and non-irrigated alfalfa hay, but the percentage difference of 74.3 percent for irrigated sorghum was even lower than for those commodities. When comparing this column to the previous table, the percentages for non-irrigated sorghum and irrigated alfalfa hay were lower than expected. This may further support the hypothesis that producers changed non-irrigated to irrigated acres for those commodities.

Caution must be used in examining the other four column categories in Table 3. To explain, when an FSA value is twice the reported value, the percentage difference is not equal to the absolute percentage difference when a reported value is twice the FSA value. For example, when the FSA value is 100 and the reported value is 50, the percentage difference is 25 percent (recall 100 is added to each data source), but the absolute percentage difference is 33.3 percent when the values are reversed. However, it is worth mentioning two commodities, irrigated sorghum and proso millet, had the highest percentage differences in the extreme columns (i.e., <-50% and >50%). Furthermore, it is interesting to add the two columns displaying the positive and negative percentage differences and compare them to Table 2. For example, for non-oil sunflowers, 12 percent of the reports had positive percentage differences and 6 percent had negative percentage differences. Although twice the reports had positive percentage differences, the percentage difference was negative at the macro-level. Therefore, the dispersion of data for reporting units with negative percentages (i.e., the reported value was greater than the FSA value) was much greater than for units with positive percentages. In addition to non-oil sunflowers, this phenomenon also occurred for irrigated corn, irrigated sorghum, and oil sunflowers. Irrigated soybeans was the only commodity where the reverse relationship happened (i.e., the number of reporting units with negative percentage differences was higher, but the percentage difference was positive at the macro-level). Discussion about chickpeas is being omitted because the number of reports is too small.

5. Modifications to Methodology After First Year of Implementation

An analysis of the survey indications in December 2006 and March 2007 showed mixed results pertaining to the precision of the acreage indications for both major commodities, such as corn and soybeans, and minor commodities, such as dry beans and sunflowers. Both the precision and accuracy of the survey indications for on-farm grain stocks were extremely unfavorable (i.e., the CVs were very high and the indications were much too low). In addition, using the FSA list frame, a valid zero for planted crop acres could not be differentiated from missing data. To explain, when FSA 578 administrative data were not present for a commodity, the producer could have either not grown the commodity (i.e., a valid zero) or grown the commodity but not reported it to FSA. This caused a few operational issues because survey procedures utilize valid zeros on the NASS list frame. Since any methodological changes had to be specified and implemented before the sample needed to be selected, there was not sufficient time to analyze the survey indications from June 2007 and September 2007. However, the results from December 2006 and March 2007 provided sufficient information to pursue several improvements.

First, although the FSA 578 administrative data were more current than NASS list frame control data, the geographic area being covered was significantly less, since an FSA farm was smaller in size than a NASS-defined farm. In Nebraska, on average, there were 2.4 FSA farms for every NASS-defined farm. The sample size for most quarters was increased to increase the amount of land covered. This change was accomplished by modifying the rotation scheme, as shown in Table 4. In order to have better control over the quarterly sample size, the number of replicates was changed from 3 to 7 and the same rotation approach was used for each component. In an attempt to minimize respondent burden, the total sample size for all four months was kept approximately the same as the original approach. Although about the same number of producers would be contacted, they would be contacted more often than before. The overall (four quarter) sample size was about 900 larger than that using the NASS list frame, while the quarterly sample sizes ranged from 100 smaller to 500 larger.

Table 4: Modified QAS Rotation Scheme

Sample Component	Replicate Number			
	December	March	June	September
General, Small Grains, Row Crops	1,2,3,4,5,6,7	1,2,3,6	1, 3,5,6	2,3,5,6,7

Second, on-farm grain storage capacity was added into the sampling process. This change could be implemented because NASS internally developed a file of FSA farm numbers with control data for on-farm grain storage capacity. Unfortunately, only a partial file could be created because of time constraints. This partial file was created using two sources, first of which was previously reported survey data. All FSA farms that had reported data for on-farm grain storage capacity during the QAS in December, March, or June were identified. The second source, the NASS list frame, was used to create a file of name and address information for farms with on-farm grain storage capacity. Due to timing, only farms on the NASS list frame with 5,000 or more bushels of on-farm grain storage capacity were included. This file was then matched to the FSA list frame population. For one-to-one matches, the FSA farm was assigned the control data value from the NASS list frame. When multiple FSA farms matched to one NASS farm, the FSA farms were identified on the FSA GIS data layer and the control data value from the NASS list frame was transferred over. This revised FSA GIS file was then overlaid with satellite data, and the FSA farm's on-farm grain storage capacity was estimated by visually inspecting the circular bins of all FSA farms associated with the NASS farm and assigning values which would sum to the control data value. Again, due to time constraints, the manual process focused on determining the control data values for FSA farms that had the largest NASS grain storage values associated with them. At the end of the process, FSA farms identified as having on-farm grain storage capacity from either source were matched to the new FSA list frame population and on-farm grain storage capacity was added to the other control data for matches.

Third, some producers had difficulty reporting for the parcel of land even though a map was included on the front of the questionnaire as a visual aid. This problem was mentioned by enumerators during the data collection phase. In an effort to address this issue and reduce respondent confusion, boundaries and information for the legal

description and minor civil divisions (MCDs) were added to the map. Previously, the map showed only the road names.

Lastly, the logic for utilizing the FSA 578 administrative data for various survey procedures was updated. As stated earlier, the FSA 578 administrative data file contained the current-year planted crop acreages by FSA farm number for producers who reported to FSA. In the original methodology, when administrative data were not present for a commodity, the planted crop acreage was considered missing when it could have been a valid zero (i.e., the producer did not grow the commodity). The missing planted crop acreages were replaced with a zero when it could be definitely determined that the producer did not have a particular commodity. This new logic would potentially improve various survey procedures as well as certain ratios being generated as part of the survey indications.

6. Reliability of Survey Indications

NASS has a policy and standards memorandum that documents the statistical precision of the survey indications obtained from major national probability surveys conducted by the agency. This memorandum contains the desired precision levels expressed as target coefficients of variation (CVs). The target CVs are specified for key items and vary by survey and geographic level. The factors considered in determining the target CVs are: 1) expected use of the estimate by data users, 2) impact on NASS performance measures in the strategic plan, 3) compatibility between the region/state and U.S. targets, 4) relative magnitude of the estimate, 5) achievability given agency appropriated resources, workload, response burden, and quality of sampling information, and 6) historical experience. NASS does not expect the target CVs to be met every time the survey is performed, but will take appropriate action if a target CV is not consistently met over time. Possible actions include making improvements to the frame, revising the survey design, increasing the target CV, or discontinuing publication of the survey indication.

For the QAS, state-level target CVs are specified for planted acreage and on-farm grain stocks. As stated previously, the QAS is conducted quarterly in March, June, September, and December. Planted acreage of a crop is not asked every month the survey is conducted. For example, the number of acres planted to sorghum is asked in March, June, and September where the number of acres planted to corn is asked in March, June, and December. For on-farm grain stocks, a commodity may be asked every month the survey is conducted, but a target CV may not be specified every time.

To assess the performance of using FSA administrative data versus the NASS list and area frames, the CVs for survey indications in Nebraska from June 2004 through September 2006 (i.e., prior to implementing the pilot program) were compared to the CVs from December 2006 through September 2008 (i.e., after using FSA administrative data). For each time period, the percentage of instances where the target CV was met was calculated. Table 5 shows the number and percentage of target CVs met by commodity for planted acreage and on-farm grain stocks.

Table 5: Comparison of Target CVs Achieved

Commodity	Using NASS List and Area Frames (6/04-9/06)			Using FSA Administrative Data (12/06-9/08)		
	# of Target CVs met	Total #	% of Target CVs met	# of Target CVs met	Total #	% of Target CVs met
Planted Acres:						
All Wheat	8	8	100%	6	6	100%
Corn	4	7	57%	2	6	33%
Soybeans	4	7	57%	3	6	50%
Oats	8	8	100%	3	6	50%
Sorghum	7	7	100%	3	6	50%
Potatoes	0	3	0%	1	2	50%
Dry Beans	2	7	29%	1	6	17%
Sunflowers	2	7	29%	2	6	33%
On-Farm Grain Stocks:						
All Wheat	6	10	60%	0	8	0%
Corn	6	10	60%	0	8	0%
Soybeans	5	10	50%	0	8	0%
Oats	9	10	90%	3	8	38%
Sorghum	2	7	29%	0	6	0%

Using the NASS list and area frames, the target CVs were achieved a higher percentage of the time for nearly every item. There were only two items, planted acres of potatoes and sunflowers, where the percentage of target CVs met was higher using the FSA administrative data; and one item, all planted wheat acres, where the percentage was the same. Therefore, the methodology using the NASS list and area frames outperformed the methodology using the FSA administrative data with respect to target CVs for key items.

Another evaluation was conducted to compare the performance of using the NASS list and area frames against the FSA administrative data. Rather than examining only key items with target CVs, the CVs for all survey indications (including the direct expansions, within survey ratios, and survey-to-survey ratios) were included in the analysis. The second year of the operational pilot program (i.e., December 2007 through September 2008) was utilized since the methodology was new the first year and improvements had been made. For the NASS list and area frames, the CVs for all survey indications from the last months prior to implementing the pilot program (i.e., December 2005 through September 2006) were used in the analysis. Corresponding months within the two time periods were paired (i.e., December 2007 and 2005, March 2008 and 2006, etc.). The CV for the survey indication using the FSA administrative data was then divided by the CV for the analogous survey indication using the NASS list and area frames. When the ratio was less than 0.90, the survey indication using the FSA administrative data was considered more precise; when the ratio was greater than 1.10, the survey indication using the NASS list and area frames was considered more precise. When the ratio was between 0.90 and 1.10, neither of the survey indications was considered more precise and it was classified as a tie. Table 6 shows the results of this evaluation for acreage, yield, production, and on-farm grain stocks.

Table 6: Comparison of Reliability of Survey Indications

Commodity and Corresponding Months	# of Comparisons Made			% of Comparisons Made		
	NASS Frame More Precise	FSA Frame More Precise	Tie	NASS Frame More Precise	FSA Frame More Precise	Tie
Acreage, Yield, and Production:						
Dec 07/Dec 05	170	82	30	60.3%	29.1%	10.6%
Mar 08/Mar 06	24	15	4	55.8%	34.9%	9.3%
Jun 08/Jun 06	31	9	20	51.7%	15.0%	33.3%
Sept 08/Sept 06	34	9	7	68.0%	18.0%	14.0%
Subtotal	259	115	61	59.5%	26.4%	14.0%
On-Farm Grain Stocks:						
Dec 07/Dec 05	14	1	0	93.3%	6.7%	0.0%
Mar 08/Mar 06	19	4	1	79.2%	16.7%	4.2%
Jun 08/Jun 06	8	1	1	80.0%	10.0%	10.0%
Sept 08/Sept 06	13	1	0	92.9%	7.1%	0.0%
Subtotal	54	7	2	85.7%	11.1%	3.2%
Total:	313	122	63	62.9%	24.5%	12.7%

Using the NASS list and area frames, the survey indications were more precise a higher percentage of the time for all corresponding months. In the case of on-farm grain stocks, the percentage using the NASS list and area frames was almost always above 80 percent and the precision levels using the FSA administrative data were much inferior. Again, the analysis revealed the methodology using the NASS list and area frames outperformed the methodology using the FSA administrative data.

7. Reasons for Discontinuing the Operational Pilot Program

As stated earlier, the goals of the new methodology were to reduce respondent burden, reduce data collection costs, and improve survey indications. Thus far, this paper has concentrated on the third goal. In the ensuing paragraphs, concluding remarks will be discussed with respect to the survey indications, followed by comments based on additional analysis that was performed with respect to the other two goals.

Survey indications: In general, the survey indications for acreage, yield, and production were more precise using the NASS list and area frames. It was speculated that these results were due to the FSA sample covering less total land than the NASS sample, even though the FSA control data were more current and the total list sample size was about 8 percent larger than using the NASS frames. This hypothesis is probably true since the number of FSA farms was approximately 2.4 times larger than the number of NASS-defined farms. Also, both the precision and accuracy of the survey indications for on-farm grain stocks were extremely unfavorable when using the FSA administrative data, even after the methodological changes (which required a considerable workload increase for staff) were implemented. It was speculated that the survey indications for stocks would still be unsatisfactory even if control data for on-farm grain storage capacity were available for all FSA farms (i.e., rather than for only a subset of them). This hypothesis is probably true since, unlike the situation for a NASS-defined farm, the crops harvested from an FSA farm are not typically stored on that farm.

Respondent burden: The average time to complete an interview using each methodology was compared for corresponding quarters. There was no evidence the average interview time was shorter using the FSA administrative data than it was using the NASS list and area frames. There was also no difference in the response rates between the two approaches. When considering other issues, respondent burden actually increased using the FSA administrative data. For example, a producer was asked to complete multiple questionnaires more often because it is more common for a producer to be associated with multiple FSA farms than it is multiple NASS-defined farms. Also, some producers had difficulty reporting for a specific FSA farm, but nearly all producers recognize the land associated with their NASS-defined farm. As mentioned previously, because an FSA farm was smaller in size than a NASS-defined farm, the FSA list sample size would need to be larger in order to obtain CVs comparable to the NASS frame, thus increasing respondent burden.

Data collection costs: The average cost per sample using each methodology was compared for corresponding quarters. This cost represented the total number of hours charged by enumerators during the data collection phase. There was no evidence the average cost per sample using the FSA administrative data was less than using the NASS list and area frames. The overall cost to the agency was actually higher using the FSA administrative data, since use of the FSA frame required NASS staff to perform extensive additional work related to operational survey procedures (e.g., processing and manipulation of files, survey preparation and coordination, etc.). Thus, the increase in total staff hours resulted in a higher overall cost than using the NASS list and area frames. Over time, this increase in overhead may have been reduced as the process was refined and streamlined.

In conclusion, the analysis revealed that the objectives were not sufficiently achieved. After two years, the operational pilot program of utilizing FSA administrative data as a sampling frame in Nebraska was discontinued. However, various operational programs within NASS continue to use FSA administrative data, just not as a sampling frame.

References

Kott, P.S. and Bailey, J.T., (2000), "The Theory and Practice of Maximal Brewer Selection with Poisson PRN Sampling," Proceedings of the International Conference on Establishment Surveys II, 269-278.