

Assessment of a Review Process for the 2017 Census of Agriculture

Denise A. Abreu, Linda A. Lawson and Sonia Hickman

National Agricultural Statistics Service, 1400 Independence Ave SW,
Washington, DC 20250

Abstract

The National Agricultural Statistics Service (NASS) conducts a Census of Agriculture every 5 years, in years ending in 2 and 7. The Census uses a list frame. The 2017 Census used capture-recapture methods to produce published estimates. NASS's June Area Survey (JAS) was used as the independent survey in the capture-recapture framework. The JAS uses an area frame and tracts of land are classified as agricultural tract or non-agricultural tract based on the agricultural activity of the area. Capture-recapture requires a matched dataset consisting of all matches of a Census record to a JAS tract. The matching is performed using record linkage. Good quality name and address (N&A) information on both sources is important to the accuracy of the matching. In recent years, NASS has devoted substantial resources to train field interviewers to collect complete N&A information on the JAS non-agricultural tracts. Both in the 2012 and 2017 Censuses, an extensive review of a subset of the non-agricultural tracts was conducted prior to matching. These reviews were costly and demanded several months of Agency resources. This paper evaluates whether the 2017 review process is essential to the Census of Agriculture adjustments for undercoverage, nonresponse and misclassification.

Key Words: List frame, area frame, capture-recapture, cost analysis

1. Introduction and Background

The United States Department of Agriculture's National Agricultural Statistics Service (NASS) conducts hundreds of surveys and prepares reports that cover every aspect of U.S. agriculture. The majority of the reports provide estimates that impact U.S. markets and the prices of commodities. Some examples of these include corn, soybeans, wheat and upland cotton estimates of acreage and forecasts of yield. NASS conducts the Census of Agriculture every 5 years, in years ending in 2 and 7. The Census provides information on characteristics of U.S. farms and ranches and the people who operate them. A farm is defined to be any place from which \$1,000 or more of agricultural products were produced and sold or normally would have been sold during the year. During the census, data are collected on land use and ownership, operator characteristics, production practices, income and expenditures, and numerous other characteristics. The census provides the most uniform comprehensive agricultural data for every county in the nation. It is used by federal, state and local governments and others who provide services to farms and rural communities. Census estimates are published at the national, state and county levels. The estimates impact community planning, availability of operational loans and other funding, location and staffing of service centers, and farm programs and policies.

The Census is a list-based endeavor. The list contains both agricultural operations that are in the target population (farms) and agricultural operations that are not in the target population (non-farms). The Census Mailing List (CML) is incomplete; not all farms are on the list. To account for farming operations not on the CML, NASS uses the June Area

Survey (JAS). The JAS uses an area frame and, during pre-screening, tracts of land are classified as agricultural tract or non-agricultural tract based on the agricultural activity of the area. The JAS is conducted annually and also provides an estimate of the number of farms. In 2007, the difference in the estimated number of farms from the Census and from the JAS was larger than could be attributed to sampling error alone (Abreu et al. 2010). This led to the decision to use capture–recapture methodology as the foundation for adjusting the 2012 Census of Agriculture, and any future censuses, for undercoverage, nonresponse, and misclassification (Young et al. 2017).

Capture-recapture requires a matched dataset consisting of all matches of a Census record to a JAS tract. The matching is performed using probabilistic record linkage. The JAS tracts that do not match a Census record are referred to as the Not-on-Mail List domain (NML). Records on the CML and NML are mailed a Census questionnaire of different colors to allow identification of the two domains. Good quality name and address (N&A) information on both sources (Census and JAS) is important to the accuracy of the matching. The accuracy of the matching process is critical for producing a precise weighting adjustment.

Because the Census is a mailout/mailback data collection effort, NASS allocates substantial time and resources to improve the N&A information on the CML. Estimates derived from the JAS are based on the information collected from the agricultural tracts only. The N&A information for all tracts is obtained during the pre-screening process prior to the JAS data collection. Due to time constraints field interviewers prioritize the pre-screening of records to focus on the agricultural tracts. As a result, many of the JAS non-agricultural tracts end up with partial or incomplete N&A information.

Prior to matching the 2017 Census and 2017 JAS records, a thorough review of the JAS non-agricultural tracts N&A information was conducted in an attempt to improve the accuracy of the matching process and the precision of the estimates based on the NML component. This review process cost over \$50,000 and required several months of staff time. This paper describes the resources allocated as well as the details of the review process.

2. June Area Survey (JAS)

The JAS is conducted annually and is based upon an area frame, which ensures complete coverage of all land within the 48 coterminous United States. For each state, land within the area frame is divided into homogeneous strata based on percent cultivated land and further into substrata based on similarity of agriculture. The land within each substratum is divided into primary sampling units (PSUs). PSUs are sampled from each substratum. Then smaller, similar-sized segments of land (about one square mile) are delineated within each selected PSU. One segment is randomly sampled from each selected PSU to be fully enumerated during the JAS (See Figure 1).

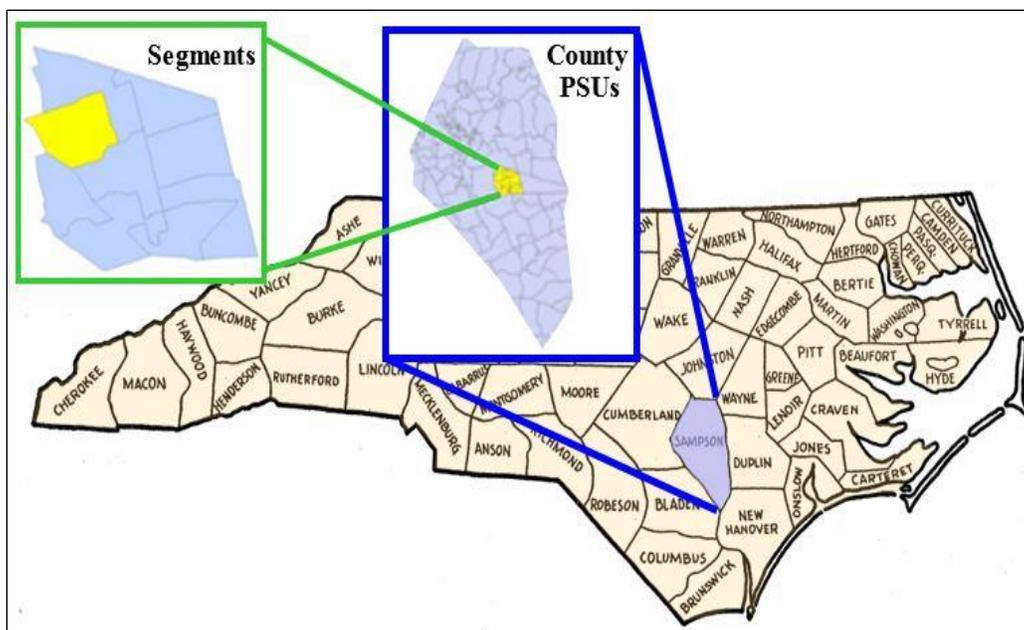
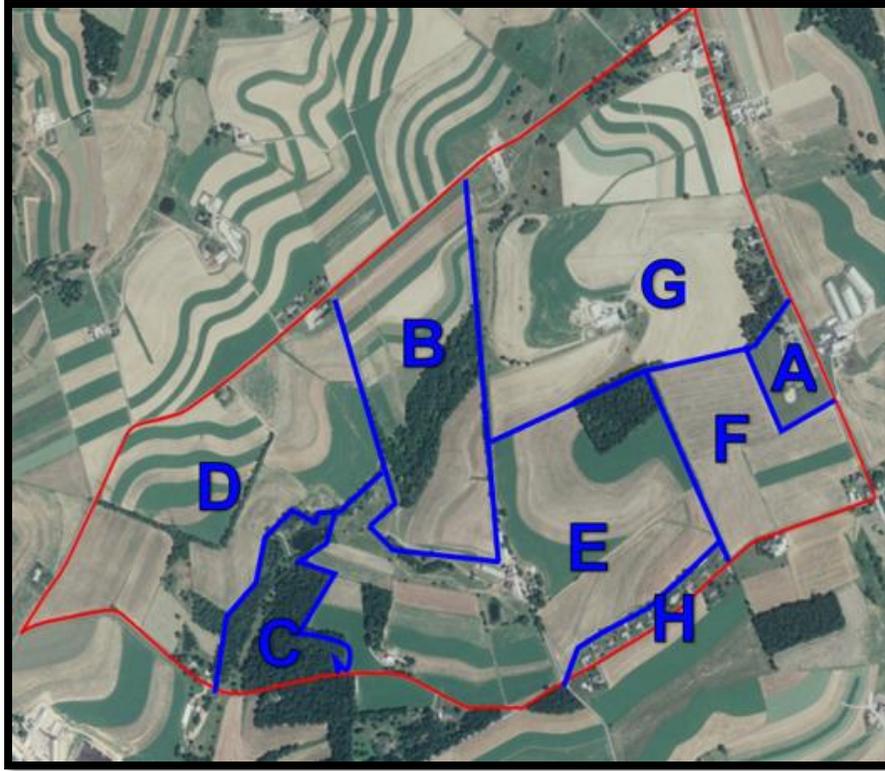


Figure 1: NASS area sampling frame for North Carolina

The JAS has a rotating panel design where twenty-percent of the sample is replaced with new segments annually and segments that have been in the sample for five years are rotated out. The newly rotated-in segments (new segments) are pre-screened in May, prior to the June data collection period, to identify segment boundaries (outlined in red in Figure 2), agricultural and non-agricultural areas within the segment, and the N&A information of possible owners and operators. Field interviewers are provided name and address information from Farm Service Agency, plat maps, and county segment maps, among other things, to help with the pre-screening. They are also instructed to conduct internet searches in their attempt to determine who operates the land. For previously enumerated segments the names and addresses are available from the previous year. Yet many of the older segments still need improvements on the N&A information.

Field interviewers are provided a paper aerial photograph showing the sampled segment area and must account for all land inside the segment boundary. They divide each segment into tracts of land (outlined in blue in Figure 2). Obvious non-agricultural areas, such as roads, rivers, etc., are assigned a tract letter and automatically classified as a non-agricultural tract. Each of the remaining tracts of land is assigned a tract letter that represent a unique land operating arrangement. These tracts are then screened for agricultural activity and classified as either an agricultural tract or a non-agricultural tract.

JAS data collection is conducted during the first two weeks of June when field interviewers return to interview only the agricultural tract farm operators. Because the primary purpose of the JAS is to provide crop and livestock acreages, field interviewers spend most of their pre-screening time on improving the information on the agricultural tracts for new segments. As a result, many of the non-agricultural tracts have partial or incomplete N&A information.



**Figure 2: The area outlined in red is the segment.
Tracts are outlined in blue and labeled with letters.**

3. Utilizing County Assessor Information to Intersect JAS Segments

In the past, NASS has purchased geo-referenced real estate property information to aid in the identification of farmers in the JAS segments (Abreu et al. 2015). In an effort to improve the quality of the N&A information of the non-agricultural tracts prior to matching to the 2017 CML, it was decided to purchase geo-referenced real estate property information from ReportAll, Inc. ReportAll is a supplier of spatial data with an interface that provides integrated access to a multitude of government record sites and other internet based mapping platforms covering more than 148.3 million parcels in over 2,953 counties in the U.S.

The JAS segments with the non-agricultural tracts that had partial or incomplete N&A information were sent to ReportAll. The cost from ReportAll to NASS was \$7,000. NASS provided ReportAll with the JAS segment boundaries in Geographic Information Systems format (See Figure 1). In turn, ReportAll provided NASS with a geo-database of property parcels. These parcels were intersected against the NASS JAS segments and clipped to only show the parcels within the JAS segments (See Figure 2).



Figure 3: Area outlined in red is a sampled JAS segment.



Figure 4: JAS sampled segment intersected with ReportALL parcels outlined in yellow.

If the parcel partially overlaid the segment, only the portion that was within the segment was displayed. In addition, ReportAll provided NASS with any property parcel information that was available, such as address, owner name, some land use information, tax ids, etc.

4. Methodology

The capture-recapture framework requires a matched dataset consisting of 2017 Census of Agriculture records overlapping 2017 JAS tracts. This will be only a subset of the census records. Capture-recapture also requires, all JAS tracts (both agricultural and non-agricultural). The 3.1M records on the CML were matched to the 119,108 JAS tracts using probabilistic record linkage.

Because field interviewers did not always obtain complete N&A information for the non-agricultural tracts, it was necessary to evaluate the quality of the JAS non-agricultural tracts prior to conducting the record linkage matching process. Table 1 shows the distribution of JAS agricultural and non-agricultural tracts.

Table 1. Distribution of JAS Agricultural and Non-agricultural Tracts

Type of Agricultural Tract	Number of Tracts	Percent
Agricultural tracts	47,442	39.8
Non-agricultural tracts	71,666	60.2
Total	119,108	100.0

Of the 71,666 non-agricultural tracts, 9,364 had questionable or partial N&A information. Upon additional analysis, these were divided into four groups based on the quality of the information contained within the records (See Table 2). The first group of 3,267 records had enough partial information to be able to do a basic search by name or property address. The second group of 1,288 records did not have enough information to perform a basic search. However, their information could be found by spatially linking them to the

ReportAll parcel records of land owners available at the county level. Over 26 percent of the non-agricultural tracts were comprised of the vacant homes, Government Land, residential areas & obvious non-agricultural areas such as roads and bodies of water. Therefore, it was determined that these 2,505 records did not require additional N&A improvements and were placed into a third group. The remaining 24.6 percent of the records were placed in a fourth group because they did not have enough information to conduct a basic search and the county was not covered by ReportAll. (No parcel listing was available).

Table 2. Non-agricultural Tracts Grouped Based on the Quality of the N&A Information

Group	Type of records	Number	Percent
1	Records had enough partial information to be able to do a basic search by name or property address.	3,267	34.9
2	Records did not have enough information to be able to do a basic search but could be spatially linked to parcel records of land owners	1,288	13.7
3	Records did not require additional N&A improvements (i.e., vacant homes, obvious non-agricultural areas such as roads)	2,505	26.7
4	Records did not enough have information to research and improve and the county was not covered by ReportAll	2,304	24.6
	Total	9,364	100.0

For purposes of matching to the CML, the focus was placed on improving the N&A information for the non-agricultural tracts in groups 1 & 2. Due to the nature of the incompleteness of the N&A information of the records in groups 1 and 2, two reviewing strategies were implemented. To prepare the data prior to the start of both review processes required 5 full-time staff for one to two weeks, at a cost of \$20,000.

5. Phase I Review – Group 1

The first review consisted of performing a basic search for the records in group 1. Fourteen staff from NASS's Data Collection Centers (DCC) were tasked with reviewing the 3,267 records with partial N&A information. DCC staff were given 3 weeks to conduct this review. This review cost \$8,000 in staff time and resources.

DCC staff were provided a spreadsheet containing the partial information for each non-agricultural tract. For each tract, there was enough information in either the name or the address fields to be able to do a basic search. For example, a name with a street, but no house address, or a complete address without a name. In addition, a column on the spreadsheet indicated whether the JAS segment was in a county covered by the ReportAll file. That did not necessarily mean that there would be any N&A information; however, it provided a search start for the DCC staff. ReportAll did not have 100 percent coverage of all the counties in the US. In addition, if a county was covered, that did not guarantee that N&A information was available for all parcel land owners. See Table 3 below for an example of the format of the spreadsheet.

Table 3. Sample Spreadsheet Provided to DCC Staff

County	Segment	Tract	Report All	Name	Address	City	State	Zip Code
County	150471	G	Yes	John Doe	Hilltop Road	Any Town	XX	99999
County	150471	J	Yes	Woods	4318 CR 17	Any Town	XX	99999
County	970126	A	No	House	123 Main Street	Any Town	XX	99999

DCC staff, first looked at the ReportALL column on the spreadsheet. If the information was not in the ReportALL. Then, staff proceeded to do lookups using the CLEAR search tool. CLEAR is a Thomson Reuters Online Investigation Software. If the information was not on CLEAR, then they would search on their state/county internet tax records, where available. Once the DCC staff identified the new information, they would proceed to update the information for that JAS non-agricultural tract.

6. Phase II Review

The second review involved the 1,288 non-agricultural tracts that did not have enough partial information to be able to do a basic search. The segments containing these non-agricultural tracts were spatially linked to the ReportALL parcel records of land owners. Figure 5 shows a JAS segment as completed by a field interviewer during the JAS data collection. The blue outlines represent the tract boundaries. In this case, non-agricultural tracts G and J have incomplete N&A information. Figure 6 shows the same JAS segment intersected with the ReportALL database.



Figure 5: Area outlined in red is a sampled JAS segment and letters G and J represent non-agricultural tracts with incomplete information.



Figure 6: JAS sampled segment intersected with ReportALL parcels outlined in yellow.

The yellow outlines represents land owner parcels from the ReportALL file. Each parcel had a listing of names associated with it. Sometimes there would be multiple parcels within a tract or the tract was larger than the ReportALL parcel. If the entire parcel corresponded to the JAS tract letter, then N&A information was updated for that non-agricultural tract. If more than 90% of the tract was associated with a parcel, then the name and address was utilized.

Four staff from the Research and Development Division (RDD) worked on reviewing and updating the N&A information for these tracts. RDD staff had to complete this review in 3 weeks. It cost \$18,000 in staff time and resources to conduct this review.

7. 2017 CML Matching Results

Upon completion of both review processes, the updated non-agricultural tracts were included as part of the probabilistic record linkage process to the 2017 Census of Agriculture CML. Table 4 describes the final outcome of each of the updated non-agricultural tracts.

Table 4. Final Outcome of Each Updated Non-Agricultural Tract

	DCC Review	Percent	RDD Review	Percent	Total	Percent
Records linked to CML record	272	8.3%	54	4.2%	326	7.2%
Records not linked but mailed NML questionnaire	2,174	66.5%	272	21.1%	2,446	53.7%
Determined to be obvious non-agricultural	821	25.1%	962	74.7%	1,783	39.1%
Total	3,267	100.0%	1,288	100.0%	4,555	100.0%

Over 60 percent of the updated non-agricultural tracts were included as part of the 2017 Census of Agriculture data collection effort. Results show that 7.2% of the updated tracts were linked to a record on the CML. While, 53.7% were mailed a census questionnaire as part of the NML domain. Almost 40% of the tracts updated were determined to be obvious non-agricultural operations (i.e., churches, golf course, Government land, etc). Fewer obvious non-agricultural tracts (25.1%) came from those tracts with partial/incomplete information (DCC review) as compared to those with little to no information available (RDD review).

8. Conclusions & Future Research

The preparation and review to improve the N&A information of the non-agricultural tracts cost NASS over \$50,000. Of the non-agricultural tracts that had their information improved, 61.1% were included in the 2017 Census of Agriculture data collection activities. These records were either part of the original CML or were mailed a questionnaire as part of the NML domain. Because the 2017 Census of Agriculture is still on-going, no further results can be provided at this time. The 2017 Census of Agriculture estimates are to be released on February 2019.

Further research should be conducted on the non-agricultural tracts that had the N&A information improved. Is important to evaluate the impact these improved records had on the number of U.S. farms. Did the allocated resources justify the gains in terms of its effect on the estimate of the number of farms in the U.S.? In addition, identifying the characteristics of these records could help determine future list building efforts for upcoming censuses and other list frame building activities for the list of farm operations NASS maintains.

9. References

Abreu, D. A., J. S. McCarthy, and L. A. Colburn (2010). Impact of the Screening Procedures of the June Area Survey on the Number of Farms Estimates. Research and Development Division. RDD Research Report #RDD-10-03. Washington, DC: USDA, National Agricultural Statistics Service.

Abreu, Denise A., Matt Deaton, and Wendy Barboza. 2015. Enhancing the June Agricultural Survey Pre-Screening through the Use of County Assessor's Information. Proceedings of the Joint Statistical Meetings, Seattle, WA.

Young, Linda J., Lamas, Andrea C. and Abreu, Denise A. "2012 Census of Agriculture: A Capture-Recapture Analysis." *Journal of Agricultural, Biological, and Environmental Statistics*. 22:523-539.