



**National
Agricultural
Statistics
Service**

Research and
Development Division
Washington DC 20250

RDD Research Report
Number RDD-11-01

October 2011

On the Feasibility of Using NASS's Sampling List Frame to Evaluate Misclassification Errors of the June Area Survey

Denise A. Abreu
Andrea C. Lamas
Hailin Sang
Kenneth K. Lopiano
Pam Arroway
Linda J. Young

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

EXECUTIVE SUMMARY

During the past three years, the National Agricultural Statistics Service (NASS) has made an effort to address, quantify and adjust for misclassification on its annual June Area Survey (JAS). Misclassification occurs (1) when an operating arrangement is identified as a non-farm but qualifying agricultural activity is present or (2) when a non-farm arrangement is incorrectly identified as a farm. A farm is defined as a place from which \$1,000 or more of agricultural products were produced and sold, or normally would have been sold, during the year, and the computation includes any government agricultural payments received. Misclassification is a direct cause of an undercount in the number of farms indication produced annually by the JAS.

Through a cooperative agreement between NASS and the National Institute of Statistical Sciences (NISS), a research team was created in 2009 to review the methodology associated with the JAS and to recommend changes that would address this undercount. In 2009, the team evaluated the use of the 2007 Census of Agriculture mailing list (CML) to assess misclassification on the 2007 JAS. The results indicated that the CML was a rich source to account for the undercount of farms on the JAS. However, the CML is only available every five years. Because the JAS is an annual survey, misclassification should be assessed each year.

NASS maintains a list of farmers and ranchers, referred to as the list frame, from which the yearly list-based surveys' samples are selected. The list frame is updated on an on-going basis and operators are categorized as either active or inactive. Active list records are those assumed to have a high likelihood of being farming operations. Inactive records are those such as deceased operators, farms no longer in business, idle facilities, landlords, etc. Many of the active records represent agricultural establishments that operate land but do not have sufficient production to be classified as a farm in a specific year. However, they are maintained on the list frame as active records to help ensure high coverage of farms for the Census of Agriculture every five years. There are also pure active status inaccuracies that exist on the list frame. That is, there are records identified as "active" that are out-of-business or no longer operate any agricultural land or facilities. The team is exploring the potential of using the list frame on a yearly basis to assess misclassification on the JAS.

After the 2007 Census of Agriculture, the farm/non-farm status of 2007 list frame records was evaluated. This analysis showed that 72 percent of the active records on the list frame were identified as farms on the census. Thus, 28 percent of the records identified as active on the list frame were actually non-farms. This indicates that the census list frame contains active records that are not associated with farms (farm status inaccuracies). A similar distribution of list frame farm status inaccuracies is anticipated for non-census years. For this work, the 2009 JAS records were matched to the 2009 list frame and to the 2009 Farm Numbers Research Project (FNRP) records. Characteristics associated with matched records that agreed and disagreed in the farm/non-farm classification were explored. Subsequent efforts will focus on whether these characteristics could be used to effectively model the probability that a list frame record is a farm.

The results showed that 2,068 list frame records matched to FNRP records. Of these, 246 represented operations incorrectly identified as farms (using an algorithm based on active status and total value of sales) on the list frame. These misclassified operations were mostly active ones, and they were spread evenly across the various cultivated strata with only a few occurring in the agri-urban or commercial strata. Half were small with less than \$10,000 in agricultural sales; 46 percent had \$10,000 to \$250,000 in sales, and the remaining operations had sales exceeding \$250,000. An additional 61 operations were identified as non-farms on the list frame but were farms according to the FNRP. These operations were located primarily on non-agricultural tracts without potential and represented marginal farms with less than \$10,000 in value of sales.

Of the 2,068 list frame operations that matched to FNRP, 1,276 had a completed FNRP interview and 792 had their FNRP data estimated. The characteristics of FNRP tracts incorrectly identified as farms on the list frame were similar for both completed and estimated FNRP surveys. The attributes of the list frame records incorrectly identified as non-farms were also similar for tracts with completed and estimated surveys.

The overall results of this study showed that the FNRP was an important tool in the identification of the list frame farm status inaccuracies and confirmed the presence of misclassification on the list frame. Therefore, if the list frame is used to adjust for misclassification on the JAS without considering its farm status inaccuracies, the JAS number of farms indication could be biased upwards.

RECOMMENDATION

1. *Research and evaluate potential ways in which the list frame's farm status inaccuracies can be reliably identified and excluded from any adjustments to the June Area Survey (JAS).* The results of this analysis confirm the presence of some list frame farm status inaccuracies. If the list frame is used to adjust for misclassification on the JAS without considering its farm status inaccuracies, the JAS number of farms indication could be biased upwards.

On the Feasibility of Using NASS's Sampling List Frame to Evaluate Misclassification Errors of the June Area Survey

Denise A. Abreu¹, Andrea C. Lamas¹, Hailin Sang², Pam Arroway³,
Kenneth K. Lopiano⁴, Linda J. Young⁴

Abstract

During the past three years, the National Agricultural Statistics Service (NASS) has made an effort to address, quantify, and adjust for an undercount in the number of farms indication from its annual June Area Survey (JAS), which is based on an area frame. This undercount is a direct result of the misclassification of agricultural tracts as non-agricultural. The 2007 Census of Agriculture mailing list (CML) was evaluated as a potential source to assess misclassification on the 2007 JAS. The CML was found to be a rich source from which to quantify the undercount of farms on the JAS. However, the CML is only available every five years, and misclassification on the JAS should be assessed each year. Independently of the area frame, NASS maintains a list of agricultural operators, referred to as the list frame. Yearly list-based samples are selected from the list frame. In addition, the list frame serves as the foundation for building the CML. The list frame is updated on an on-going basis and operators are categorized as either active or inactive. Although the CML includes all active records, some of these do not qualify as farming operations. This research report explores the potential of using the list frame on a yearly basis to assess the misclassification of farms on the JAS.

KEY WORDS: misclassification errors, area frame, list frame, record linkage, re-screening survey

¹ National Agricultural Statistics Service, USDA, 3251 Old Lee Hwy, Fairfax VA 22030

² National Institute of Statistical Sciences, 19 T.W. Alexander Drive, Research Triangle Park, NC 27709

³ Department of Statistics, North Carolina State University, Raleigh, NC 27695

⁴ Department of Statistics, University of Florida, Gainesville, FL 32611

1. INTRODUCTION

Each year the National Agricultural Statistics Service (NASS) publishes an estimate of the number of farms in the United States (U.S.) based on the June Area Survey (JAS). A farm is defined as a place from which \$1,000 or more of agricultural products were produced and sold, or normally would have been sold, during the year, and the computation includes any government agricultural payments received. An independent estimate of the number of farms is published from the quinquennial Census of Agriculture, which is conducted in years ending in 2 and 7. At the end of each five-year period, the annual estimates based on the JAS number of farms indication are adjusted based on intercensal trends. The annual estimate of the number of farms from the JAS has been declining steadily between censuses (especially between the 2002 and 2007 Censuses) as depicted in Figure 1. In 2007, the estimate from the JAS was significantly below that from the census; and the required intercensal trend adjustment to the JAS was unexpectedly large as shown by the circled area in Figure 1.

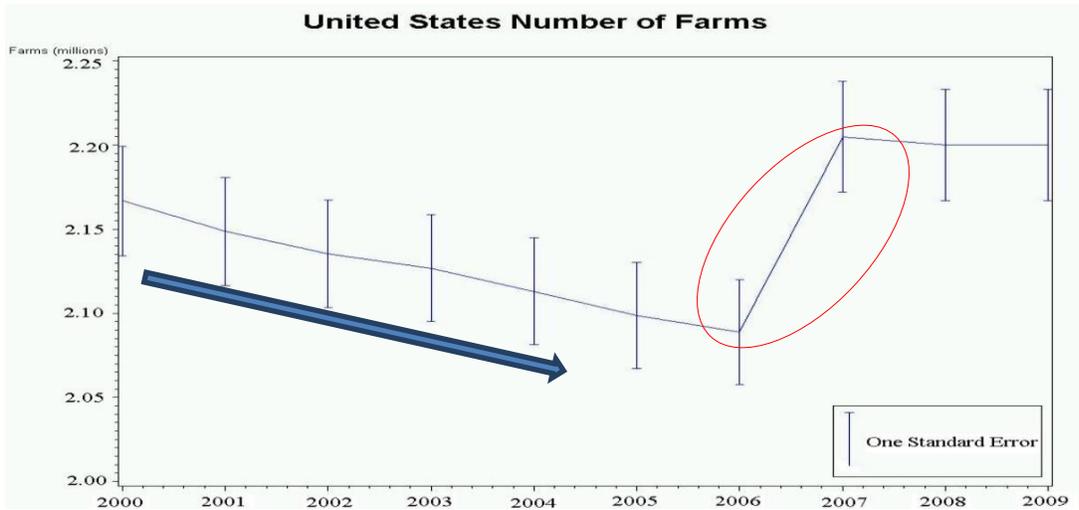


Figure 1: Published estimates of the number of U.S. farms from 2000 to 2009 and bars with a length of one standard error on either side of the estimate.

During previous studies conducted by NASS, misclassification was identified as a source of the underestimation in the JAS (Abreu 2007; Johnson 2000). Misclassification occurs (1) when an operating arrangement with qualifying agricultural activity is identified as a non-farm, or (2) when a non-farm arrangement is incorrectly identified as a farm. One study of misclassification (Abreu, Dickey and McCarthy, 2009) revealed that some agricultural operations were incorrectly classified as non-agricultural during JAS pre-screening. These results led to more intensive efforts to understand the source and extent of misclassification in the JAS so that it could be addressed. One effort was the Farm Numbers Research Project (FNRP), based on an intensive post-June survey re-screening in 2009 (Abreu, McCarthy and Colburn, 2010). Concurrently, this undercount issue was also addressed by a team of researchers formed to review the methodology associated with the JAS and to recommend changes, through a collaborative agreement with the National Institute of Statistical Sciences (NISS). This latter team consists of two NASS researchers, two university faculty members, a post doctoral fellow, and a graduate student. The team has considered several measures to address the issue of misclassification on the JAS. Through matching the JAS to the Census of Agriculture list frame, the team evaluated

misclassification on the JAS (Abreu, et al. 2010) and then developed appropriate methodology to adjust for misclassification during non-census years (Lamas, et al. 2010). In addition to misclassification, the team identified non-response as another source contributing to the JAS undercount. In Lopiano, et al. (2010), the effect of estimation of agricultural activity for some JAS sampled units is discussed, and methodology for adjusting for both non-response and misclassification is developed. Because the census is only conducted every fifth year, the team further proposed a yearly follow-on survey to the JAS called the Annual Land Utilization Survey (ALUS) (Arroway et al. 2010). ALUS would make the JAS a two-phase sample. In addition to providing information about misclassification of farms and non-farms, it would allow for proper assessment of misclassification and result in an improvement in all JAS indications. However, because ALUS would lead to greater costs associated with the JAS, alternative methods that would not require enumerators to collect further data are attractive. One possibility is to use NASS's annual list frame to assess misclassification in the JAS during non-census years. The team's current effort is focused on evaluating the potential for this approach. The initial results are discussed in this report.

2. THE JUNE AREA SURVEY (JAS)

The June Area Survey (JAS) is based on an area-frame and collects information about U.S. crops, livestock, grain storage capacity, and type and size of farms. The distribution of crops and livestock can vary considerably within each state in the United States. Therefore, the precision of the survey indications can be substantially improved by dividing the land within each state into homogeneous groups (strata) and optimally allocating the total sample to the strata. The basic stratification employed by NASS involves: (1) dividing the land into land-use strata such as intensively cultivated land, urban areas and range land, and (2) further dividing each land-use stratum into substrata by grouping areas that are agriculturally similar. The JAS uses a sample comprised of designated land areas (segments) selected from this stratification. A typical segment is about one square mile (i.e., 640 acres). Each segment is outlined on an aerial photo that is provided to the appropriate field enumerator (See red outlined area in Figure 2).

Through field enumeration, a segment is divided into tracts of land, each representing a unique land operating arrangement (Refer to blue outlined areas in Figure 2). An area screening form that provides an inventory of all tracts within the segment and contains screening questions that determine whether or not each tract has agricultural activity is completed for all sample segments. Using this form, all land inside the segment is screened for agricultural activity, and the screening applies to all land in the identified operating arrangement (both inside and outside the segment). Those operations (tracts) that qualify as agricultural are subsequently interviewed using the area version questionnaire, which collects detailed agricultural information about the operator's land, again both inside and outside the segment.



Figure 2: JAS Segment (outlined in red) and Tract Boundaries (outlined in blue)

The area frame is a theoretically complete sampling frame with every acre of land having a known probability of selection. As such, it is used to estimate the number of farms and land in farms independently of the list frame. The area frame also provides a measure of incompleteness in the list. The JAS uses a replicated sample design. A sample rotation scheme is used to reduce respondent burden caused by repeated interviewing and to avoid the expense of selecting a completely new area sample each year. Once selected, a segment stays in the sample for five years, so that annually, approximately 20 percent of the sampled segments in each land-use stratum are replaced with newly selected segments in each land-use stratum. Full descriptions of the JAS design and analysis procedures may be found in Davies (2009) and Lamas et al. (2011), respectively.

3. THE NASS LIST FRAME

Each year, NASS conducts hundreds of list-based surveys. The agency maintains a list of farmers and ranchers from which the samples for these list-based surveys are selected. This list frame also serves as the foundation for the development of the Census Mail List (CML). NASS builds and improves the list on an ongoing basis by obtaining outside source lists. Sources include lists from state and federal government agencies, producer associations, seed growers, pesticide applicators, veterinarians, marketing associations, and a variety of other agricultural sources. NASS also obtains special commodity lists to address specific list deficiencies. These outside source lists are matched to the NASS list using record linkage programs. Most names on newly acquired lists are already on the NASS list. Records not on the NASS list are treated as potential farms until NASS can confirm their existence as a qualifying farm. Each operation on the list frame is categorized as active or inactive. Active list records are assumed to have a high probability of representing active farming operations. Inactive list records may be associated with landlords, deceased operators, farms no longer in business, etc. Many of the active records represent agricultural establishments that operate land but do not have sufficient production to be

classified as a farm in a specific year. However, they are maintained on the list frame as active records to help ensure high coverage of farms for the Census of Agriculture every five years. There are also pure active status inaccuracies that exist on the list frame. That is, there are records identified as "active" that are out-of-business or no longer operate any agricultural land or facilities.

The question being considered here is whether the NASS list frame can be used to assess misclassification in the JAS in non-census years. After the 2007 Census of Agriculture, the farm/non-farm status of 2007 list frame records was evaluated. Seventy-two percent of the active list frame records matched to farms on the census. The remaining 28 percent were found to be non-farms,⁵ indicating that the census list frame contains active records that are not associated with farming operations (farm status inaccuracies). If the list frame farm status inaccuracies are not considered when adjusting for misclassification on the JAS, the adjustment for misclassification will be larger than it should be. Thus, for the list frame to be useful in assessing misclassification in the JAS, a method of properly accounting for the list frame farm status inaccuracies must be developed.

4. THE FARM NUMBERS RESEARCH PROJECT (FNRP)

In 2009, NASS conducted the Farm Numbers Research Project (FNRP). FNRP was a one-time follow-on survey to the 2009 JAS for the first-year rotation segments (Abreu, McCarthy and Colburn, 2010). Recall the design of the JAS includes rotating in new segments each year, with segments staying in the JAS sample for five years. Each year's sample is comprised of segments from each of five rotations. Thus, the 2009 JAS contained segments that were rotated into the sample in 2009, 2008, 2007, 2006 and 2005. The FNRP targeted the 20 percent of JAS segments that were newly rotated in for 2009 ("2009 segments"). For the FNRP, all tracts in the 2009 segments that were non-agricultural or estimated in JAS were revisited. In our present framework, FNRP information could subsequently be used to verify the farm/non-farm status of the 2009 list frame records. That is, FNRP provides the "gold standard" on farm status for 2009 list frame records with the limitation that the FNRP only constituted 20 percent of the 2009 JAS.

5. MATCHING 2009 JAS TO THE 2009 LIST FRAME

Probabilistic record linkage was used to match all 2009 JAS agricultural and non-agricultural tracts to the 2009 list frame records in 42 states. The analysis excluded the New England states because those files were not available at the time the match was processed. The JAS is only conducted in Hawaii during census years, and Alaska does not have an area frame. Records were brought together into link groups, each of which possibly represented the same operation. Routinely, link groups are classified into one of three distinct types: definite match, possible match or non-match (Broadbent et. al., 1999). Possible matches are identified for Field Office (FO) staff to review. However, in the interest of saving time and resources, no FO review was conducted. Instead, only two distinct types of matches were identified: match and non-match. Eliminating the FO review from the linkage process led to a more conservative approach in identification of matches and non-matches. That is, to maximize the quality of the final results,

⁵ Internal analysis conducted by Thomas Jacob of the Information Management Group.

all possible matches were treated as non-matches. Consequently, some true matches became non-matches.

From the 83,203 original JAS tracts, 92,152 names and addresses were identified. These were prepared and standardized for matching to the list frame. For this linkage, all agricultural and non-agricultural tracts were considered. Partner records and records with additional information were also included for each JAS tract to maximize matching results. In addition to the name and address information, existing area-to-area and area-to-list links were used to bring records together. After each June survey, FOs conduct a yearly overlap/non-overlap process in which JAS agricultural tracts are overlapped to the list frame. This provides a measure of list incompleteness. JAS identification numbers (IDs) are stored for each list frame record overlapped to the JAS (area-to-list links). In addition, an unduplication of the area frame records is conducted. The ID of any area record matching another area record (area-to-area links) is stored. These identification numbers were used during matching to bring records together that would not have come together solely based on name and address information.

From the 2009 list frame, 4,683,345 names and addresses were prepared and standardized for matching to the 2009 JAS. This list included both active and inactive records. Every year, certain records are purged from the list frame, usually because they have been inactive for more than five years. The only records excluded were those flagged to be purged from the list frame due to extended inactivity.

When matching, the ideal scenario is to have one area record match one list record. However, after the initial matching, some link groups had more than one tract and others had more than one list frame record. Although the area file was set up to have only one tract per link group, in some cases, more than one tract occurred in a link group, indicating that different tracts matched to the same list records. To address this issue, tracts were split into separate groups and all list records that matched were assigned to both split groups. When multiple list records matched one tract, the list frame records were ranked and based on their active/inactive status, the “best” one was selected using the following rules:

Ranks Used to Assign the Best of Several List Frame Records to a JAS Tract

Rank	List Record Type	Description
1	Active target	Assumed to be farming operations
2	Potential CML	Non-respondents to any of the agricultural surveys conducted routinely to update active status of the list frame
3	Active partner	Partners associated with active target
4	Inactive	Deceased operators, farms no longer in business, idle facilities, landlords, etc.
5	Other	Hired managers, etc.

6. RESULTS

The results of the linkage yielded 41,926 matches. Table 1 shows the breakdown of the matched tracts by type of agricultural tract as identified in the JAS. The vast majority of the matches were to agricultural tracts (86.4 percent). This is not surprising because the list frame is targeted for agricultural operations, and agricultural tracts have the most complete name and address information. During JAS screening procedures, non-agricultural tracts are classified into the following three types: potential for agriculture unknown, having potential for agriculture, and not having potential for agriculture. Non-agricultural tracts without potential comprised almost 12 percent of all the matches.

Table 1. Matched JAS and List Frame Records by Type of Agriculture as Identified by the JAS

Type of Agricultural Tract	Number Tracts Matched	Percent
Agricultural tracts	36,245	86.4
Non-agricultural tracts w/ potential	546	1.3
Non-agricultural tracts w/ unknown potential	240	0.6
Non-agricultural tracts w/out potential	4,895	11.7
Totals	41,926	100

Further evaluating the matches by the rank, Table 2 shows that 87.6 percent of the matches were to active list frame records. Additionally, 3.1 percent were matches to potential CML records, and 9.1 percent were to inactive list frame records. Routine follow-up for the list frame does not include the inactive list frame records so their farm status will be difficult to determine.

Table 2. Results of “Best” List Records Matched to Area Tracts

List Record Rank	Total Tracts Records	Percent
Active target - 1	36,725	87.6
Potential CML - 2	1,320	3.1
Active partner - 3	27	0.1
Inactive - 4	3,831	9.1
Other -5	23	0.1
Totals	41,926	100.0

The results of the initial matching of the area tracts and/or list frame records are displayed in Table 3 below. Notice that, as described earlier, some link groups had more than one list frame record and/or more than one JAS tract. Only link groups with one JAS tract matching to the list frame records are considered for evaluating the farm/non-farm status of the list frame records. In addition, any tract matching 10 or more list frame records was excluded from further analysis. Thus, the highlighted cells in Table 3 represent the 36,439 matched records that are considered here (23,951 were 1-to-1 matches, 12,488 were 1-to-many). Again, for the link groups with 2 or more list frame records, only the best ranked one was selected. All results that follow focus on the 36,439 matches.

Table 3. Numbers of Link Groups for Each Combination of Number of JAS Tracts and Number of List Frame Records Within a Link Group⁶

Tracts in Link Group	Number of List Records per Link Group										Total
	1	2	3	4	5	6	7	8	9	10+	
1	23,951	6,412	3,008	1,401	784	426	230	130	97	224	36,663
2		1,364		1,028		694		436		836	4,358
3			96			111			92	305	604
4+				19				40		242	301
Total	23,951	7,776	3,104	2,448	784	1,231	230	606	189	1,607	41,926

7. ASSESSING THE LIST FRAME’S FARM/NON-FARM STATUS

Each JAS agricultural tract was identified as a farm or non-farm in June based on whether it had \$1,000 in sales of agricultural products or 1,000 points based on the potential for agricultural products produced (if sales were less than \$1,000), and the computation includes any government agricultural payments received. All non-agricultural tracts were considered non-farms. Identifying farms on the list frame is important because list frame records lack a farm/non-farm status. Subject-matter experts from the List Frame Section recommended assigning farm status based on the active status code and total value of sales. This approach was adopted and led to the following algorithm for assigning farm status to the list frame records:

- 1) For active records (AS = 0) records, an operation is identified as a farm if sales exceeded \$1000; otherwise it is taken to be a non-farm.
- 2) All inactive records (AS = 1 - 8, 10 - 19) are identified as non-farms based on enumerator information, regardless of any sales value on the records.
- 3) The census-only records (AS = 9) are generally present so that farms in multiple counties or multiple states are represented for census purposes, but not for the survey program. These are not considered further.
- 4) The farm status of the records assigned AS = 30-31 and 33– 36 are unknown.
- 5) Operations with AS = 32 are receiving CRP payments and are considered farms.
- 6) An operation headquartered out of state is assigned AS = 40 and is assumed to be a farm.
- 7) An operation that has a major name change (AS = 41) is assumed to be a non-farm.

Farm status was assigned to each JAS tract matched to a list record. In Table 4, the number of list frame farms, non-farms and unknowns by their agricultural/non-agricultural status on the JAS are shown. Recall that the farm status for the list frame records with active status code 30-31 and 33-36 is not known (“List Farm Status Unknown” column in Table 4 below). Utilizing the sales class information to assign farm status would have resulted in 808, or about 68 percent, of the 1,185 records being identified as having some agricultural activity. Because the proportion of these records that represent farms may differ from the proportion in the other list frame records, they are treated separately from the other matched records. The farm/non-farm status of the list frame records by rank is displayed in Table 5 below. The farm/non-farm status of matched records for both the list frame and the JAS is shown in Table 6 below.

⁶ All results that follow will be focused on the highlighted cells in Table 3.

Table 4. Farm Status Assignment for 2009 List Frame Records

Type of Agricultural Tract	List Non-Farm	List Farm	List Farm Status Unknown	Total
Agricultural tracts	1,911	28,626	808	31,395
Non-agricultural tracts w/ potential	195	281	39	515
Non-agricultural tracts w/ unknown potential	73	128	20	221
Non-agricultural tracts w/out potential	2,108	1,882	318	4,308
Totals	4,287	30,967	1,185	36,439

Table 5. Farm/Non-farm Status of List Frame Records by Rank

Rank	List Non-Farm	List Farm	List Farm Status Unknown	Total
Active target – 1	814	30,711	0	31,525
Potential CML - 2	0	69	1,185	1,254
Active partner – 3	0	24	0	24
Inactive – 4	3,473	156	0	3629
Other -5	0	7	0	7
Total	4,287	30,967	1,185	36,439

Table 6. Farm/Non-farm Status of Matched List Frame and JAS Records

Farm/Non-farm Status	List Non-Farm	List Farm	List Farm Status Unknown	Total
JAS Non-Farm	2,729	2,694	464	5,887
JAS Farm	1,558	28,273	721	30,552
Total	4,287	30,967	1,185	36,439

7.1 Assessing the Accuracy of Farm/Non-farm Status Based on the List Frame

As noted earlier, about 28 percent of the 2007 list frame records identified as active were not farms for the 2007 Census of Agriculture⁵. Thus, the farm/non-farm status of the 2009 list frame records had some misclassification. If the list frame is to be used to assess misclassification in the JAS, then being able to identify the list frame farm status inaccuracies is important. The results of the FNRP, discussed earlier, are used here to provide insight into the types of tracts that are misclassified as farms or non-farms on the list frame. Because the FNRP only constituted 20 percent of the 2009 JAS, this part of the analysis is limited.

Current NASS procedures define a tract as a unique land operating arrangement; however, for densely populated tracts, multiple operations (places of interest) may have been erroneously included for any particular tract during the JAS survey enumeration. For the FNRP, the concept of subtracts was introduced to address tracts that had multiple places of interest. For a selected tract, all places of interest were considered subtracts. For enumeration purposes, if eight or more subtracts were present within a tract, these subtracts were sub-sampled at pre-determined rates.

The FNRP sample consisted of 10,204 JAS tracts, which resulted in a total of 17,191 subtracts. Only 2,226 (about 13 percent) FNRP subtracts matched to the list frame. Of these, 2,068 had only one subtract in the JAS tract. The remaining 158 subtracts were from 52 JAS tracts. These 158 subtracts are not addressed further here because visual inspection is required to link them to their corresponding list frame record.

7.2 Evaluation of FNRP Records and Their Status on the List Frame

Of the 2,068 matching records, 483 were list non-farms, 1,458 list farms, and 127 list unknowns (i.e., AS=30-31, 33-36) (Table 7). As mentioned earlier, the FNRP results are the gold standard and assumed to be accurate. Of the 683 tracts identified as non-farms by both JAS and FNRP, 246 were farms on the list frame. Similarly, 61 operations were incorrectly classified as non-farms on the list frame and the JAS. Both of these are the list frame farm status inaccuracies and confirm the presence of misclassification on the list frame.

Table 7. A Comparison of Farm/Non-Farm Status on 2009 JAS, 2009 List Frame, and FNRP

		List non-farm	List farm	List Farm Status Unknown	Total
JAS non-farm	FNRP non-farm	356	246	81	683
	FNRP farm	61	188	12	261
JAS farm	FNRP non-farm	18	25	1	44
	FNRP farm	48	999	33	1,080
Total		483	1,458	127	2,068

Ninety-eight percent of the 246 operations inaccurately identified as farms on the list frame were identified as active operations (rank 1). However, 77 percent were associated with JAS non-agricultural tracts without potential for agriculture. Only a few of the 246 were from agricultural tracts. Most of the misclassified operations are in the moderately to highly cultivated strata; very few of the misclassified operations were in the agri-urban or commercial strata. Even though half of these misclassified operations were small with less than \$10,000 in sales, over 46 percent of the remaining operations were in the \$10,000-\$250,000 sales categories. See all related tables in Appendix A.

Of the 61 operations that were non-farms on the JAS and the list frame but identified as farms by FNRP, most were marginal farms with less than \$10,000 in value of sales, and they were primarily from JAS non-agricultural tracts without potential. See related tables in Appendix B.

When evaluating the characteristics of the 188 operations that were JAS non-farms, but identified as farms on both FNRP and the list frame, it is clear that these are correctly classified on the list frame. Nearly all of them came from JAS tracts identified as non-agricultural without potential, and they were in the moderately to highly cultivated strata. See all related tables in Appendix C. Because value of sales was available for both FNRP and list frame farms, agreement of farming operations from both sources was evaluated. The highlighted cells in Table 8 below are where the list sales class and FNRP sales class agree; the two sources only agree about a third of the time. Most of the time the list frame value of sales is higher than that

reported in FNRP, indicating that the list frame is over-estimating sales and thus categorizing operations as having more sales than was reported in FNRP. It is important to note that the current list frame procedures assign value of sales based on the largest reported values. Thus, it is not surprising that it overstates sales when compared to FNRP. It does point out that value of sales alone should not be used to determine farm/non-farm status on the list frame.

Table 8. A Comparison of Sales Class Values for Matched FNRP and List Frame Records with Highlighted Cells Indicating Agreement Between the Two Sources.⁷

List Sales Class	FNRP Sales Class												Total
	\$1-\$999	\$1,000-\$2,499	\$2,500-\$4,999	\$5,000-\$9,999	\$10,000-\$24,999	\$25,000-\$49,999	\$50,000-\$99,999	\$100,000-\$249,999	\$250,000-\$499,999	\$500,000-\$999,999	\$1M-\$2.5M	\$5M+	
\$1-\$999	0	1	0	0	0	0	0	0	0	0	0	0	1
\$1,000-\$2,499	0	20	2	2	1	1	0	0	0	0	0	0	26
\$2,500-\$4,999	0	10	3	3	4	0	1	0	0	0	0	0	21
\$5,000-\$9,999	0	5	3	6	4	1	1	0	1	0	0	0	21
\$10,000-\$24,999	0	10	8	6	10	1	1	0	0	1	0	0	37
\$25,000-\$49,999	0	3	7	7	8	5	3	1	0	0	0	0	34
\$50,000-\$99,999	0	2	1	2	3	3	3	1	0	0	0	0	15
\$100,000-\$249,999	0	1	1	1	2	1	2	8	1	0	0	0	17
\$250,000-\$499,999	0	1	0	0	1	0	0	1	2	0	0	0	5
\$500,000-\$999,999	0	0	0	0	0	1	0	1	1	1	2	1	7
\$1M-\$2.5M	0	0	0	0	0	0	1	0	0	1	2	0	4
Total	0	53	25	27	33	13	12	12	5	3	4	1	188

7.3 Evaluation of FNRP Completed Interviews and Their Status on the List Frame

The 2,068 FNRP records that matched the list frame were comprised of 1,276 completed interviews and 792 estimated interviews. In this and the next subsection, the characteristics of these two groups and how they differ are explored. For the completed interviews, 199 of the operations identified as farms on the list frame were non-farms both in FNRP and the JAS (Table 9). Similarly, 61 operations were incorrectly classified as non-farms on the list frame and the JAS.

Table 9. A Comparison of Farm/Non-Farm Status on 2009 JAS, 2009 List Frame, and FNRP for Completed FNRP Interviews

		List non-farm	List farm	List Farm Status Unknown	Total
JAS non-farm	FNRP non-farm	314	199	65	578
	FNRP farm	61	174	11	246
JAS farm	FNRP non-farm	14	21	1	36
	FNRP farm	22	379	15	416
Total		411	773	92	1,276

The 199 farming operations that were inaccurately identified as farms on the list frame were primarily considered active operations (rank 1), and most of them (80 percent) were non-agricultural tracts without potential for agriculture. The misclassified operations were evenly

⁷ A \$5M row is not present for this table because there were not any list frame records with sales exceeding \$5M that matched to a FNRP record. The bars displayed on the table represent the percent contribution of the cell to the column total.

spread over the cultivated strata with very few in the agri-urban or commercial strata. Over half of these operations were small with less than \$10,000 in sales, 44 percent of the operations indicated sales of \$10,000-\$250,000, and a few indicated sales in excess of \$250,000. See all related tables in Appendix D.

All 61 operations that were non-farms on the JAS and the list frame but were identified as farms on FNRP had completed interviews during the FNRP (see Appendix B for characteristics). Similarly, the characteristics of the 174 completed interviews that were considered JAS non-farms but identified as farms on both FNRP and the list frame were the same as those observed for all 188 records (see Appendix E). The pattern of agreement of sales from both sources (Table 10) was also the same as for the overall group.

Table 10. A Comparison of Sales Class Values for Matched FNRP and List Frame Records with Highlighted Cells Indicating Agreement Between the Two Sources.⁷

List Sales Class	FNRP Sales Class												Total
	\$1-\$999	\$1,000-\$2,499	\$2,500-\$4,999	\$5,000-\$9,999	\$10,000-\$24,999	\$25,000-\$49,999	\$50,000-\$99,999	\$100,000-\$249,999	\$250,000-\$499,999	\$500,000-\$999,999	\$1M-\$2.5M	\$5M+	
\$1-\$999	0	1	0	0	0	0	0	0	0	0	0	0	1
\$1,000-\$2,499	0	20	2	2	1	1	0	0	0	0	0	0	26
\$2,500-\$4,999	0	9	2	2	4	0	1	0	0	0	0	0	18
\$5,000-\$9,999	0	5	3	4	4	1	1	0	1	0	0	0	19
\$10,000-\$24,999	0	10	8	6	10	1	1	0	0	1	0	0	37
\$25,000-\$49,999	0	3	7	7	7	5	3	1	0	0	0	0	33
\$50,000-\$99,999	0	2	1	1	2	3	2	1	0	0	0	0	12
\$100,000-\$249,999	0	1	1	1	2	0	2	6	1	0	0	0	14
\$250,000-\$499,999	0	1	0	0	0	0	0	1	1	0	0	0	3
\$500,000-\$999,999	0	0	0	0	0	1	0	1	1	1	2	1	7
\$1M-\$2.5M	0	0	0	0	0	0	1	0	0	1	2	0	4
Total	0	52	24	23	30	12	11	10	4	3	4	1	174

7.4 Evaluation of FNRP Estimated Interviews and Their Status on the List Frame

Of the 792 estimated FNRP interviews, 47 of the operations identified as farms on the list frame were non-farms both in FNRP and the JAS (Table 11). These are additional records that are likely to be incorrectly identified as farms on the list frame. In contrast, no operations were incorrectly classified as a non-farm on either the list frame or the JAS.

Table 11. A Comparison of Farm/Non-Farm Status on 2009 JAS, 2009 List Frame, and FNRP Based on Estimated FNRP Interviews

		List non-farm	List farm	List Farm Status Unknown	Total
JAS non-farm	FNRP non-farm	42	47	16	105
	FNRP farm	0	14	1	15
JAS farm	FNRP non-farm	4	4	0	8
	FNRP farm	26	620	18	664
Total		72	685	35	792

The 47 list frame farm operations that were non-farms both in FNRP and the JAS possessed similar characteristics to those with completed interviews. See all related tables in Appendix F.

The 14 estimated interviews that were JAS non-farms but identified as farms on both FNRP and the list frame were correctly treated as farms. They shared the same attributes of those with completed interviews. See all related tables in Appendix G. The highlighted cells in Table 12 correspond to agreement in the sales class for the list frame and FNRP. For operations in this group, the two sources agreed half the time. Even though the numbers are too sparse to make any firm generalizations, the pattern seems to be similar to that of the completed interviews; that is, the list frame value of sales is usually higher than what was reported in FNRP indicating that the list frame is over-estimating sales, which is not surprising.

Table 12. A Comparison of Sales Class Values for Matched FNRP and List Frame Records with Highlighted Cells Indicating Agreement Between the Two Sources.⁷

List Sales Class	FNRP Sales Class								Total
	\$1,000-\$2,499	\$2,500-\$4,999	\$5,000-\$9,999	\$10,000-\$24,999	\$25,000-\$49,999	\$50,000-\$99,999	\$100,000-\$249,999	\$250,000-\$499,999	
\$1,000-\$2,499	0	0	0	0	0	0	0	0	0
\$2,500-\$4,999	1	1	1	0	0	0	0	0	3
\$5,000-\$9,999	0	0	2	0	0	0	0	0	2
\$10,000-\$24,999	0	0	0	0	0	0	0	0	0
\$25,000-\$49,999	0	0	0	1	0	0	0	0	1
\$50,000-\$99,999	0	0	1	1	0	1	0	0	3
\$100,000-\$249,999	0	0	0	0	1	0	2	0	3
\$250,000-\$499,999	0	0	0	1	0	0	0	1	2
Total	1	1	4	3	1	1	2	1	14

8. CONCLUSIONS AND DISCUSSION

Using the FNRP as the “gold standard” to accurately identify list frame farm status inaccuracies was informative. Even though the analysis was split by type of FNRP interview (completed or estimated), the results showed to be similar for both groups whenever misclassification was present. The list frame value of sales is usually higher than what was reported in FNRP indicating that the list frame is overestimating sales. Since the current list frame procedures assign value of sales based on the largest reported values, sales will tend to be overstated when compared to FNRP. Therefore, value of sales alone should not be used to determine farm/non-farm status of records on the list frame. The farm status inaccuracies are an issue that needs to be addressed further if the list frame is used to adjust for misclassification on the JAS. The JAS number of farms indication could become biased upwards. Thus, the potential for using the list frame for misclassification adjustment of the number of farms indication depends on whether or not the list frame farm status inaccuracies can be reliably identified and excluded from the adjustment, and this merits further research.

9. RECOMMENDATION

1. *Research and evaluate potential ways in which the list frame’s farm status inaccuracies can be reliably identified and excluded from any adjustments to the June Area Survey (JAS).* The results of this analysis confirm the presence of some

list frame farm status inaccuracies. If the list frame is used to adjust for misclassification on the JAS without considering its farm status inaccuracies, the JAS number of farms indication could be biased upwards.

10. REFERENCES

Abreu, Denise A., Pam Arroway, Andrea C. Lamas, Kenneth K. Lopiano, and Linda J. Young (2010). Using the Census of Agriculture List Frame to Assess Misclassification in the June Area Survey. Proceedings of the 2010 Joint Statistical Meetings.

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, D. A., N. Dickey and J. McCarthy (2009). 2007 Classification Error Survey for the United States Census of Agriculture. RDD Research Report # RDD-09-03. Washington, DC:USDA, National Agricultural Statistics Service.

Abreu, D. A. (2007). Results from the 2002 Classification Error Study. Research and Development Division. RDD Research Report #RDD-07-03. Washington, DC:USDA, National Agricultural Statistics Service.

Arroway, Pam, Denise A. Abreu, Andrea C. Lamas, Kenneth K. Lopiano, and Linda J. Young (2010). An Alternate Approach to Assessing Misclassification in JAS. Proceedings of the 2010 Joint Statistical Meetings.

Broadbent, K and Iwig, W. (1999), "Record Linkage at NASS Using Automatch". 1999 FCSM Research Conference, <http://www.fcsm.gov/99papers/broadbent.pdf>

Davies, Carrie (2009). Area Frame Design for Agricultural Surveys. Research and Development Division. RDD Research Report #RDD-09-06. Washington, DC: USDA, National Agricultural Statistics Service.

Johnson, J.V. (2000). Agricultural Census Classification Error Estimation Using an Area Frame Approach. Data Quality Research Section. Unpublished Manuscript. Washington, DC: National Agricultural Statistics Service, USDA.

Lamas, Andrea C., Denise A. Abreu, Pam Arroway, Andrea C. Lamas, Kenneth K. Lopiano, and Linda J. Young (2010). Modeling Misclassification in the June Area Survey. Proceedings of the 2010 Joint Statistical Meetings.

Lopiano, Kenneth K., Denise A. Abreu, Pam Arroway, Andrea C. Lamas, and Linda J. Young (2010). Modeling Misclassification in the June Area Survey. Proceedings of the 2010 Joint Statistical Meetings.

Young, Linda J., Denise A. Abreu, Pam Arroway, Andrea C. Lamas, and Kenneth K. Lopiano (2010). Precise Estimates of the Number of Farms in the United States. Proceedings of the 2010 Joint Statistical Meetings.

Characteristics of the 246 operations that were farms on the list frame but non-farms on both the JAS and FNRP.

Table A1: Breakdown by Rank on List Frame Records

Rank	Number of Tracts	Percent
Active target - 1	241	98.0
Potential CML - 2	2	0.8
Inactive - 4	3	1.2
Total	246	100.0

Table A2: Breakdown by Value of Sales on the List Frame

Total Value of Sales	Number of Tracts	Percent
Less than \$999	1	0.4
\$1,000-\$2,499	42	17.1
\$2,500-\$4,999	36	14.6
\$5,000-\$9,999	46	18.7
\$10,000-\$24,999	45	18.3
\$25,000-\$49,999	24	9.8
\$50,000-\$99,999	21	8.5
\$100,000-\$249,999	22	8.9
\$250,000-\$499,999	3	1.2
\$500,000-\$999,999	5	2.0
\$1M-\$2.5M	1	0.4
Total	246	100.0

Table A3: Breakdown by Type of Agricultural Tract

Type of Agricultural Tract	Number of Tracts	Percent of Total
Agricultural tracts	7	2.9
Non-agricultural tracts w/ potential	33	13.4
Non-agricultural tracts w/ unknown potential	16	6.5
Non-agricultural tracts w/out potential	190	77.2
Totals	246	100.0

Appendix A

Table A4: Breakdown by Strata

Strata	Number of Tracts	Percent
50% + cultivated	88	35.8
15-49% cultivated	73	29.7
Agri-urban/ Commercial	7	2.8
<15% cultivated	78	31.7
Total	246	100.0

Table A5: Breakdown by Mode of Collection

Collection Mode (Code)	Number of Tracts	Percent
Known Zeroes (0)	9	3.7
Mail (1)	56	22.8
Telephone (2)	71	28.9
Face-to-Face (3)	84	34.1
CATI (4)	11	4.5
Other (19)	15	6.0
Total	246	100.0

Characteristics of the 61 operations that were non-farms on the JAS and list frame but identified as farms on FNRP

Table B1: Breakdown by Rank on List Frame Records

Rank	Number of Tracts	Percent
Active target - 1	8	13.1
Potential CML - 2	0	0.0
Inactive - 4	53	86.9
Total	61	100.0

Table B2: Breakdown by Value of Sales on the List Frame

Sales Class	Number of Tracts	Percent
\$1,000-\$2,499	18	29.5
\$2,500-\$4,999	8	13.1
\$5,000-\$9,999	19	31.2
\$10,000-\$24,999	10	16.4
\$25,000-\$49,999	2	3.3
\$50,000-\$99,999	1	1.6
\$100,000-\$249,999	2	3.3
\$250,000-\$499,999	1	1.6
Total	61	100.0

Table B3: Breakdown by Type of Agricultural Tract

Type of Agricultural Tract	Number of Tracts	Percent
Agricultural tracts	2	3.3
Non-agricultural tracts w/ potential	6	9.8
Non-agricultural tracts w/ unknown potential	1	1.6
Non-agricultural tracts w/out potential	52	85.3
Totals	61	100.0

Appendix B

Table B4: Breakdown by Strata

Strata	Number of Tracts	Percent
50% + cultivated	16	26.2
15-49% cultivated	28	45.9
Agri-urban/ Commercial	0	0.0
<15% cultivated	17	27.9
Total	61	100.0

Table B5: Breakdown by Mode of Collection

Collection Mode (Code)	Number of Tracts	Percent
Mail (1)	19	31.2
Telephone (2)	16	26.2
Face-to-Face (3)	16	26.2
CATI (4)	9	14.8
Other (19)	1	1.6
Total	61	100.0

Characteristics of 188 operations that were JAS non-farms but identified as farms on both FNRP and the list frame

Table C1: Breakdown by Rank on List Frame Records

Rank	Number of Tracts	Percent
Active target - 1	184	97.9
Potential CML - 2	1	0.5
Inactive - 4	3	1.6
Total	188	100.0

Table C2: Breakdown by Strata

Strata	Number of Tracts	Percent
50% + cultivated	73	38.8
15-49% cultivated	52	27.7
Agri-urban/ Commercial	4	2.1
<15% cultivated	59	31.4
Total	188	100.0

Table C3: Breakdown by Type of Agricultural Tract

Type of Agricultural Tract	Number of Tracts	Percent
Agricultural tracts	10	5.3
Non-agricultural tracts w/ potential	22	11.7
Non-agricultural tracts w/ unknown potential	14	7.5
Non-agricultural tracts w/out potential	142	75.5
Totals	188	100.0

FNRP Completed Interviews -- Characteristics of the 199 operations that were farms on the list frame but non-farms on both the JAS and FNRP.

Table D1: Breakdown by Rank on List Frame Records

Rank	Number of Tracts	Percent
Active target – 1	199	98.0
Potential CML – 2	1	0.5
Inactive – 4	3	1.5
Total	199	100.0

Table D2: Breakdown by Value of Sales on the List Frame

Total Value of Sales	Number of Tracts	Percent
Less than \$999	1	0.5
\$1,000-\$2,499	36	18.1
\$2,500-\$4,999	27	13.6
\$5,000-\$9,999	41	20.6
\$10,000-\$24,999	36	18.1
\$25,000-\$49,999	24	12.1
\$50,000-\$99,999	13	6.5
\$100,000-\$249,999	14	7.0
\$250,000-\$499,999	2	1.0
\$500,000-\$999,999	4	2.0
\$1M-\$2.5M	1	0.5
Total	199	100.0

Table D3: Breakdown by Type of Agricultural Tract

Type of Agricultural Tract	Number of Tracts	Percent
Agricultural tracts	4	2.0
Non-agricultural tracts w/ potential	26	13.1
Non-agricultural tracts w/ unknown potential	9	4.5
Non-agricultural tracts w/out potential	160	80.4
Totals	199	100.0

Appendix D

Table D4: Breakdown by Strata

Strata	Number of Tracts	Percent
50% + cultivated	74	37.1
15-49% cultivated	60	30.2
Agri-urban/ Commercial	5	2.5
<15% cultivated	60	30.2
Total	199	100.0

Table D5: Breakdown by Mode of Collection

Collection Mode (Code)	Number of Tracts	Percent
Known Zeroes (0)	9	4.5
Mail (1)	50	25.1
Telephone (2)	64	32.2
Face-to-Face (3)	68	34.2
CATI (4)	9	4.5
Other (19)	1	0.5
Total	199	100.0

FNRP Completed Interviews -- Characteristics of 174 operations that were JAS non-farms but identified as farms on both FNRP and the list frame

Table E1: Breakdown by Rank on List Frame Records

Rank	Number of Tracts	Percent
Active target - 1	170	97.7
Potential CML - 2	1	0.6
Inactive - 4	3	1.7
Total	174	100.0

Table E2: Breakdown by Strata

Strata	Number of Tracts	Percent
50% + cultivated	65	37.4
15-49% cultivated	51	29.3
Agri-urban/ Commercial	3	1.7
<15% cultivated	55	31.6
Total	174	100.0

Table E3: Breakdown by Type of Agricultural Tract

Type of Agricultural Tract	Number of Tracts	Percent
Agricultural tracts	9	5.2
Non-agricultural tracts w/ potential	21	12.0
Non-agricultural tracts w/ unknown potential	13	7.5
Non-agricultural tracts w/out potential	131	75.3
Totals	174	100.0

FNRP Estimated Interviews -- Characteristics of the 47 operations that were farms on the list frame but non-farms on both the JAS and FNRP.

Table F1: Breakdown by Rank on List Frame Records

Rank	Number of Tracts	Percent
Active target – 1	46	97.9
Potential CML – 2	1	2.1
Inactive – 4	0	0.0
Total	47	100.0

Table F2: Breakdown by Value of Sales on the List Frame

Total Value of Sales	Number of Tracts	Percent
\$1,000-\$2,499	6	12.8
\$2,500-\$4,999	9	19.2
\$5,000-\$9,999	5	10.6
\$10,000-\$24,999	9	19.2
\$25,000-\$49,999	0	0.0
\$50,000-\$99,999	8	17.0
\$100,000-\$249,999	8	17.0
\$250,000-\$499,999	1	2.1
\$500,000-\$999,999	1	2.1
Total	47	100.0

Table F3: Breakdown by Type of Agricultural Tract

Type of Agricultural Tract	Number of Tracts	Percent
Agricultural tracts	3	6.4
Non-agricultural tracts w/ potential	7	14.9
Non-agricultural tracts w/ unknown potential	7	14.9
Non-agricultural tracts w/out potential	30	63.8
Totals	47	100.0

Appendix F

Table F4: Breakdown by Strata

Strata	Number of Tracts	Percent
50% + cultivated	14	29.8
15-49% cultivated	13	27.7
Agri-urban/ Commercial	2	4.2
<15% cultivated	18	38.3
Total	47	100.0

Table F5: Breakdown by Mode of Collection

Collection Mode (Code)	Number of Tracts	Percent
Known Zeroes (0)	2	4.2
Mail (1)	6	12.8
Telephone (2)	7	14.9
Face-to-Face (3)	16	34.0
CATI (4)	2	4.2
Other (19)	14	29.8
Total	47	100.0

FNRP Estimated Interviews -- Characteristics of 14 operations that were JAS non-farms but identified as farms on both FNRP and the list frame

Table G1: Breakdown by Rank on List Frame Records

Rank	Number of Tracts	Percent
Active target - 1	14	100.0
Total	14	100.0

Table G2: Breakdown by Strata

Strata	Number of Tracts	Percent
50% + cultivated	8	57.1
15-49% cultivated	1	7.1
Agri-urban/ Commercial	1	7.1
<15% cultivated	4	28.6
Total	14	100.0

Table G3: Breakdown by Type of Agricultural Tract

Type of Agricultural Tract	Number of Tracts	Percent
Agricultural tracts	1	7.1
Non-agricultural tracts w/ potential	1	7.1
Non-agricultural tracts w/ unknown potential	1	7.1
Non-agricultural tracts w/out potential	11	78.6
Totals	14	100.0