



**National
Agricultural
Statistics
Service**

Research and
Development Division
Washington DC 20250

RDD Research Report
Number RDD-09-03

September 2009

2007 Classification Error Survey for the United States Census of Agriculture

Denise A. Abreu
Nancy J. Dickey
Jaki S. McCarthy

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

In censuses of agriculture prior to 2007, classification errors were measured by comparing an operation's status on the census to its status on the Area Frame based June Agricultural Survey. In cases where there were discrepancies between the two, the June Agricultural Survey was assumed to be correct (since personal enumeration is used), and the operations were counted as a misclassification on the census. More benefits can be gained from examining why errors occur rather than estimating the amount of classification error, since classification error is not used to adjust census numbers.

For 2007, the primary purpose of the Classification Error Survey was to identify reasons for discrepancies between the June Agricultural Survey and the census. The 2007 Classification Error Survey was a qualitative examination of why classification and reporting errors occur. Many operations do not report consistently between the June Agricultural Survey and the Census of Agriculture and the reasons for these discrepancies cannot be determined from the questionnaires alone. Discrepancies may be due to legitimate changes in acres operated between June and the end of the year; misclassification of the operation in the June Agricultural Survey or the census; or some other reporting error.

For the 2007 Classification Error Survey, June Agricultural Survey reports for respondents in Arizona, Georgia, Minnesota, New York, and Washington were matched to their 2007 Census of Agriculture reports. In cases where acreage or operating status did not agree, respondents were provided with copies of their June Agricultural Survey and census questionnaires and reinterviewed to determine the reasons for the discrepancies. Additionally, respondents were asked some follow-up questions about reporting total acres operated.

The 2007 Classification Error Survey results showed that most of the discrepancies were actual errors which occurred in the June Agricultural Survey, not the census. Very few of the discrepancies observed between the two surveys represented real changes in the farming operation. These results also suggested that screening methods for the June Agricultural Survey should be reviewed. There were numerous cases of operations incorrectly classified as non-agricultural that had non-agricultural land inside the segment, but additional agricultural land outside the tract. The survey also found that proxy respondents tend to provide inaccurate information in the June Agricultural Survey. More efforts should be made to contact the primary operator and not proxy respondents in June.

The 2007 Classification Error Survey showed that classification error in the census remains minimal and is probably smaller than previous estimates. It is recommended that future Classification Error Survey endeavors should not focus on estimating misclassification since it is so small, especially in terms of net impact on the estimates. If there are changes made to procedures in the June Agricultural Survey or Census of Agriculture, these can be evaluated as part of the content test for the next census.

RECOMMENDATIONS

- 1. Improve data collection and classification of operations in the June Agricultural Survey.** Efforts to measure the amount of misclassification in the June Agricultural Survey are already underway. In addition to that research, the use of proxy respondents should be minimized and enumerator training on procedures for proper screening of operations should also be emphasized. If procedures are altered, the next Census Content Test should be used to evaluate them. A sample of respondents from the 2010 JAS should be included in the Census Content Test in December 2010. The data from the two reports should be compared as has been done in the 2007 CES.
- 2. Do not use future Classification Error Surveys to estimate misclassification.** The amount of misclassification in the census is small, especially in terms of net effect on estimates, and the 2007 CES showed that it is much less than would have been estimated using the JAS assumption of truth. More can be gained by examining the reasons for reporting errors in a Classification Error Survey than from attempting to estimate classification error.
- 3. Consider using information about problems with reporting in the Land Section of the census in any redesign of the census form.** Most of the acreage problems within the census were related to the Land Section. This section should be the focus of future content tests.

2007 Classification Error Survey for the United States Census of Agriculture

Denise A. Abreu
Nancy J. Dickey
Jaki S. McCarthy¹

Abstract

Following the 2002 Census of Agriculture, a Classification Error Study was conducted to estimate the number of operations misclassified (either as farms or non-farms) in the census. This was done by matching operations who reported in the Area Frame June 2002 Agricultural Survey to their census report and comparing their answers. The information on the June Agricultural Survey was assumed to be correct, since it was collected in person by trained enumerators, while the census data were obtained in most cases through a self-administered mail form. Misclassification estimates for 2002 were generated based on cases where the census report was classified differently than the matching June Agricultural Survey report. The estimated misclassification rate was small but it was clear that, in some cases, the assumption of the June Agricultural Survey response being correct was not justified.

Since the 2002 misclassification estimates were not used to adjust published census estimates, a different approach was taken for the classification error study in 2007. For 2007, the focus was on understanding why operations reported differently in the June Agricultural Survey than they did on the census, rather than on estimating misclassification rates. Census records were matched to operations' reports from the 2007 June Agricultural Survey, but neither report was assumed to be "the truth." As in 2002, this study targeted operations classified as farms in one case and non-farms in the other. In addition, it focused on operations who reported total acres operated that differed by more than 25 percent between June and the census. Instead of assuming one source was correct, these operations were reinterviewed, shown their June Agricultural Survey and census questionnaires, and asked to resolve and explain the discrepancies. In addition, operators were asked general questions related to suspected problems in reporting their acreage.

The reinterviews uncovered several different sources of errors in reporting. These occurred in both the June Agricultural Survey and the census, with the majority of errors in the June Agricultural Survey. Errors were related to respondents, enumerators and National Agricultural Statistics Service procedures and show that a multi-part solution is needed to address them.

KEY WORDS: Reporting Errors, Classification Errors, Reinterviews, Record Linkage, Reconciliation

¹ Denise A. Abreu, Nancy J. Dickey and Jaki S. McCarthy are Statisticians with the National Agricultural Statistics Service - Research & Development Division, located at Room 305, 3251 Old Lee Highway, Fairfax, VA 22030. A special thanks to Mark Apodaca, Kara Daniel, Amanda Dawson, Diane Murphy, Carol Pinto, and Angela Umlauf for their role in making this research project a success.

1. HISTORICAL PERSPECTIVE

The Census of Agriculture, which is conducted by the National Agricultural Statistics Service (NASS), is conducted every five years (for years ending in 2 and 7) and is a complete count of United States (U.S.) farms and ranches and the people who operate them. The census collects data on land use and ownership, operator characteristics, production practices, income and expenditures, and many other characteristics. The outcome, when compared to earlier censuses, helps to measure trends and new developments in the agricultural sector of our nation's economy. The information is used only for statistical purposes and data are published only in tabulated totals. 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 census year. Census forms are sent to all known and potential agricultural operations in the U.S. The census provides the only source of uniform, comprehensive agricultural data for every county in the nation.

In addition to a Census of Agriculture, NASS conducts an Area Frame based survey which collects information about U.S. crops, livestock, grain storage capacity, and type and size of farms. The June Agricultural Survey (JAS) uses a sample comprised of designated land areas (segments), which field enumerators visit to collect data on all agricultural activity occurring within the segments. A typical segment is about one square mile, which is equivalent to 640 acres. Each segment is outlined on an aerial photo which is provided to the appropriate field enumerator. The Area Frame is a theoretically complete sampling frame with every acre of land having a known chance of selection. As such, it can be used both as a stand-alone frame and to measure incompleteness in the list. A segment is divided into tracts of land, each representing a unique land operating arrangement. Crops and land located within the boundaries of a tract are associated with the tract. An Area Screening Form, which inventories all tracts within the segment and contains screening questions that determine if the tract qualifies as a farm, is completed for all segments. All land inside the segment must be screened for agricultural activity and the screening applies to all land in the identified operating arrangement, including land outside the segment. Those who qualify are interviewed using the Area Version questionnaire, which collects information specifically about the operator's land, both inside and outside the segment.

Following each census, an evaluation is conducted to measure misclassification of farms on the census mail list. Each record on the census is either in-scope (IS), i.e. a farm, or out-of-scope (OS) i.e. a non-farm. Classification errors on the census consist of undercount due to farms incorrectly classified as non-farms, overcount due to non-farms incorrectly classified as farms, and overcount due to duplicate farms.

Prior to and including the 1997 Census of Agriculture, a list-based reinterview sample of census respondents in a Classification Error Survey (CES) was used to measure classification errors on the census. Separately, the NASS Area Frame served to measure incompleteness of the census mail list, which is by far the largest component of coverage inaccuracies on the census.

Following the 1997 Census, a real-time study known as the Classification Error Study was conducted for the eleven western states of Arizona, California, Colorado, Idaho, Montana,

Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming, which comprise the West Census Region. This was done to evaluate the feasibility of using the NASS Area Frame both to measure misclassification errors and to replace the Classification Error Survey reinterview approach that was being used. The 1997 Classification Error Study results indicated a net undercount of 27,971 farms for those 11 states. While the standard error of this estimate is not available to determine statistical significance, even if statistically significant, it represents a relatively small portion of the overall number of farms. Recommendations were to replace the Classification Error Survey (reinterview approach) with the Classification Error Study (Area Frame approach) (Johnson, 2000).

After the 2002 Census of Agriculture, the Classification Error Study using an Area Frame approach was conducted in the 48 conterminous states. This was a quantitative study and its main objective was to determine whether the relative size and likelihood of classification errors warranted future studies. The underlying basis for the analysis and all quantitative results was to assume the Area Frame information represented truth. Census records were matched to Area Frame records and differences in scope of the operations between the two surveys were identified. Results of the 2002 study showed a net misclassification overcount of 51,345 farms for the U.S. The classification error was small and was not used to adjust census numbers. The eleven states from the 1997 study were also compared in 2002 and results indicated a statistically insignificant net overcount of 5,438 farms.

The results of the 2002 study indicated that although the Classification Error Study comprised a small portion of the overall coverage number, it needed to be addressed further. The 2002 Classification Error Study found an overall misclassification overcount of farms, while in the 1997 Classification Error Study, there was a net misclassification undercount of farms. Due to this inconsistency from census to census, it was recommended to conduct the study again in 2007 with a focus on addressing reasons for discrepancies between June and the census instead of a quantitative measure of the errors (Abreu, 2007).

2. BACKGROUND OF THE 2007 CLASSIFICATION ERROR STUDY (CES)

For any given year, the farm versus non-farm classification for operations should generally agree between the JAS and the census for that year. Real changes in farm statuses across these two survey collections do occur, but are generally rare. However, many operations do not report consistently between them and the reasons for these discrepancies cannot be determined from the questionnaires alone. Some discrepancies are due to legitimate changes in acres operated between June and the end of the year, though again these are relatively rare. More frequently, the discrepancies are the result of misclassification of the operation in either the JAS or the census. This may be due to how the forms were processed, because it was not clear what should be reported on a questionnaire, because only part of the operation was included, or due to some other reporting error.

Because classification error is not used to adjust census numbers, more benefits can be gained from examining why errors occur rather than from estimating the amount of classification error.

For 2007, the primary purpose of the CES was to identify reasons for the discrepancies between the two sources, both true and spurious differences. This type of analysis was suggested in an independent evaluation of the census by the 2007 Council on Food, Agriculture, and Resource Economics (C-FARE). In regard to possible classification errors, the panel suggested investigating potential coverage issues arising from new (birth) and exiting (death) farms following the 2007 JAS but before the 2007 Census of Agriculture. The frequency of such births is one of the issues that was examined in the CES.

The C-FARE panel also made two important recommendations relating to edit and imputation. First, they recommended that NASS analyze the incidence of data imputation for each question with the goal of improving the quality of the questionnaires for future Censuses of Agriculture. A large number of imputations for a question would suggest that wording needs to be improved or the question eliminated. Secondly, they pointed out that the agency should analyze the edits made to the 2007 Census with a goal of improving the quality of the 2012 Census. Again, an excessive number of edits to a question would suggest the question needs to be examined for improved wording.

There are two questions (item K46 and item K798) on the census which ask Total Acres. From the 2002 Census, 34 percent of the reported item K46 (Total Acres of Land in this Place) data were edited. The second 'Total Acres' question (item K798, which is derived by summing acres by land use) was edited 31 percent of the time. The 2005 Census Content Test (CCT) also showed problems in reporting for these items. K46 did not equal K798 in 32 percent of the cases reported (McCarthy, 2007), although it is stated clearly on the form that answers to the two questions should be the same. These questions appear essentially the same way on the 2007 Census form as they did on the 2002 and 2005 CCT forms; therefore, one would anticipate a similar level of edits for these items. For the sections where those questions appear, the CES provides the opportunity to suggest different wording and format changes that may improve the quality of the data and/or reduce analyst review and editing.

In the 2007 CES, discrepancies were examined between the 2007 Census of Agriculture and the 2007 JAS (including the data collected on segments added for the 2007 Agricultural Coverage Evaluation Survey). The CES was a qualitative examination of why errors occur, both classification and reporting errors. For this endeavor, it was necessary to provide respondents with their previously reported data from both the 2007 Census of Agriculture and the 2007 JAS. In cases where they did not agree, we tried to determine the reason by reinterviewing the respondents. Reinterviews have been used by NASS to examine potential problems in survey reporting (Hanuschak et. al., 1991). Reconciliation interviews at other organizations have also proved extremely beneficial in identifying reasons for discrepancies in terms of comprehension, recall, encoding, response options, or other problems (Morton et. al., 2008). In the CES reinterviews, respondents were asked to review their questionnaires and resolve discrepancies, if possible. They were then asked some follow-up questions about reporting total acres operated. The census collects acreage data for 2007 with no specific reference to time. The JAS, on the other hand, asks for data as of June 1st. The focus of the re-contacts was on operations with large acreage discrepancies between June 2007 and the census, and those classified as in-scope on one and out-of-scope on the other.

The objectives of this study were to examine: 1) if the change in acreage was legitimate; 2) if respondents were reporting incorrectly; and 3) if the forms were processed correctly in both cases. If the intent of the forms and how to report is unclear to respondents, improvements to these forms and processing procedures for the next census may be necessary. The information could help improve the quality of the data and/or reduce analyst review and editing.

3. METHODOLOGY

For the 2007 CES, additional name, address, and telephone information was collected on both the 2007 JAS and the 2007 Agricultural Coverage Evaluation Survey (ACES) through the addition of three questions to the survey instruments. These questions collected information on landlords, additional addresses, and names (i.e., spouse, partners) which could be related to the operation. States were selected based on a number of factors, one of which was the relative number and magnitude of discrepancies they showed in the 2002 CES. States with substantial misclassification issues in 2002 were considered strong candidates for 2007. States involved in the 1997 study were also considered, since their past experience would be helpful and their inclusion would provide more opportunity to track the results over time. An evaluation of workloads based on prior record linkage projects was also considered in state selection, since knowledge and expertise of Field Office (FO) staff plays an important role in arriving at sound and viable reasons for reporting errors. Since the study would not be conducted in all states, fair representation of the various census regions was also important to be able to make any generalizations of the results. Considering all these factors, Arizona (AZ), Georgia (GA), Minnesota (MN), New York (NY), and Washington (WA) were chosen for the 2007 CES. The Colorado (CO) and NY FOs assisted in testing procedures and instructions.

Probabilistic record linkage was used to match this additional information to the names and addresses on the 2007 Census Mail List (CML) for the five states. Probabilistic record linkage is a technique used to identify records that were believed to correspond to a CML record. Records were brought together into link groups which possibly represented the same operation. Each link group was classified into one of three distinct types: definite match, possible match or non-match (Broadbent et. al., 1999). Definite matches consisted of record pairs that with great certainty identified the same operation. Non-matches were singleton area records which did not match any census record. Possible matches were record pairs which required manual review by FO staff for determination of match/non-match status. Non-matches were considered out of scope for the study.

The linkage process matched 317,763 2007 CML records to 14,498 names and addresses from the 2007 JAS and ACES. FO staff reviewed the possible matches in their state to determine match vs. non-match status. Additional non-matches were then excluded from the study. There were 5,086 link groups prior to FO review. Table 1 shows the CES possible matches reviewed by each state participating in the survey.

Table 1: 2007 Record Linkage Review Workloads by State

State	JAS Names	Census Records	Possible Matches	Definite Matches	Total Link Groups
AZ	1,495	29,612	332	145	477
GA	3,190	82,191	749	494	1,243
MN	4,501	98,717	1,125	857	1,982
NY	1,334	52,326	149	286	435
WA	3,978	54,917	588	361	949
TOTAL	14,498	317,763	2,943	2,143	5,086

Only the possible and definite matched records remained in the scope of the 2007 CES. FO review resulted in 3,694 link groups eligible for selection. These link groups were divided into three groups based on specific characteristics of the JAS and census records and the action they would require. The groups identified were: 1) Classification in agreement (census and JAS both in-scope (IS), or census out-of-scope (OS) and JAS non-agricultural) with comparable acres; 2) Classification in agreement, with acreage differences more than 25 percent; and 3) Classification conflicts (census in-scope and JAS non-agricultural, or census out-of-scope and JAS in-scope). Table 2 presents the breakdown of the records by reinterview status. The table shows the general group description, detailed characteristics of the records, the action that was undertaken, the total number of records in each group, that group number's percentage of the total in all groups, and the number of records in each group targeted for reinterview.

Table 2: Identifying Groups to Reinterview

Group Description	Characteristics of records	Action	Total	Percent	Targeted Reinterviews
Classification in agreement, acres comparable	Census IS / JAS IS OR Census OS / JAS Non-agricultural	No Action	1,629	44.4	
Classification in agreement, acres not w/in 25%	Census IS / JAS estimated IS	No Action; assume JAS incorrect	240	6.5	
	Census IS / JAS IS	Reinterview	1,122	30.6	193
Classification Conflict	Census OS by NASS / JAS IS	FO Review Only	158	4.3	
	Census OS / JAS IS	Reinterview	185	5.1	19
	Census OS / JAS estimated IS	No Action; assume JAS incorrect	53	1.5	
	Census IS / JAS Non-agricultural	Reinterview	279	7.6	88
	Total		3,666 ²	100	300

² There were 28 census respondents identified as deceased. These were not eligible for a re-interview and were excluded from all future counts.

Based on the characteristics of the records within each of the three groups, operations eligible for a reinterview were identified. No reinterview was necessary for records where the census and JAS were both correctly scoped (either as farms or non-farms) and their acreages were comparable (within 25 percent). The groups identified to be reinterviewed were:

- 1) Acreage differences: Census in-scope and JAS in-scope records with acreage differences more than 25 percent;
- 2) Scoping differences: Census records out-of-scope and JAS in-scope; and
- 3) Scoping differences: Census in-scope and JAS out-of-scope.

Operations whose census record was determined to be out-of-scope by NASS, but whose JAS record was classified as IS were manually reviewed by FO staff to address the differences in the reports. No action was taken on any records where the JAS data were estimated. In these cases, the JAS was assumed to be incorrect.

CES interviews began in July 2008. Because we were asking respondents to reconcile data reported in June 2007, respondents who were contacted for the 2008 JAS were excluded from the scope of the CES to avoid any time period confusion. The 2008 JAS refusals and inaccessibles were also excluded. The number of respondents in each category targeted for reinterview is shown in Table 2.

In order to examine discrepancies between the 2007 Census of Agriculture and the 2007 JAS, a reinterview questionnaire was developed with probing questions. Prior to utilizing the questionnaire, it was pretested on nine respondents in Virginia.

For each contact, field enumerators received a packet containing a copy of the respondent's 2007 JAS questionnaire, a copy of the respondents' 2007 Census of Agriculture Report Form, and a 2007 Classification Error Survey Questionnaire. Enumerators reviewed the data on the operations' questionnaires before they conducted the CES interviews. The enumerators needed to know what had been reported and to be aware of any other information on either the JAS or census forms. If it was obvious why there was a discrepancy (for example, the same data were reported on both forms, but NASS classified them differently), the operation was not reinterviewed.

The CES questionnaire was completed through face-to-face interviewing as the respondent needed to be able to review their JAS and census forms. Showing the respondents their questionnaires helped refresh their memories as there was a substantial time lag between the CES and when the census and JAS had been conducted. It was important that the person who completed the questionnaires was the one reinterviewed for the CES. During the interviews, respondents were asked to review their questionnaires and determine which figure (census or JAS) was correct and to explain the discrepancy. They were also asked to provide detailed comments on the nature of the reporting differences and on some general follow-up questions about reporting acreage.

Interviews were conducted between July 7 and August 15, 2008. After the reinterviews were completed, enumerators returned the questionnaires to the FOs. The FO staff keyed the information into a reinterview data entry application, which was developed as part of the census system processing tools. Table 3 shows the number of completed reinterviews by state.

Table 3: Targeted Reinterviews by Scoping and Acreage Criteria

State	Scoping Differences	Acreage Differences > 25-percent	Totals
AZ	0	5	5
GA	32	38	70
MN	14	41	55
NY	7	31	38
WA	14	32	46
Total	67	147	214 ³

4. RESULTS

Table 4 presents the total number of completed review and reinterview cases, across all states. It shows the general group description, detailed characteristics of the records, the action undertaken, the total number and percentage of records in each group, and the number of records in each group which were reviewed by FO staff or reinterviewed by an enumerator. The table also presents the number of operations which were part of the 2007 JAS, which would have been eligible for a reinterview but were instead resolved in the field office.

³ Of the 300 cases initially considered for a re-interview, FO staff determined that 86 cases could be resolved in the office without the need to re-interview the respondents. The evaluation of these records revealed that procedural errors had been made during the census edit processing system. Several of these cases were incorrectly classified as in-scope by the edit system; however, their census reports were clearly identified as non-farms. In a number of the reports, enumerators made observations and/or did not understand the area interviewing process. Also, a number of cases involved duplication of census reports. Finally, some of the JAS reports were matched to census replicated records for which an actual census response was unavailable.

Table 4: Total Number of Completed Review & Reinterview Cases

Group Description	Characteristics of records	Action	Total	Percent	Completed Review & Reinterview Cases
Classification in agreement, acres comparable	Census IS / JAS IS OR Census OS / JAS Non-agricultural	No Action	1,629	44.4	
Classification in agreement, acres not w/in 25%	Census IS / JAS estimated IS	No Action; assume JAS incorrect	240	6.5	
	Census IS / JAS IS	Reinterview – Acreage discrepancies	586	16.0	147
	Census IS / JAS IS / Multiple Census Reports	FO Review Only – Duplicate Reports	536	14.6	154
Classification Conflict	Census OS by NASS / JAS IS	FO Review Only – NASS out-of-scope	158	4.3	44
	Census OS / JAS IS	Reinterview – Scoping discrepancies	19	0.5	9
	Census OS / 2008 JAS IS / 2007 JAS	No Reinterview	166	4.5	
	Census OS / JAS estimated IS	No Action; assume JAS incorrect	53	1.4	
	Census IS / JAS Non-agricultural	Reinterview – Scoping discrepancies	88	2.4	58
	Census IS / 2008 JAS Non-agricultural / 2007 JAS	No Reinterview	191	5.2	
	Total			3,666 ²	100

4.1 Evaluation of Classification Errors

4.1a Review of NASS Out-of-scope & Duplicate Reports

An additional set of records were not reinterviewed but required review to assess the discrepancies between the JAS and the census. These included census records which were made out-of-scope by the computer process or classified by data analysts as not qualifying as farms; and cases where two or more census records matched a single area record and at least one of them had acreage differing by more than 25 percent.

During the census review process, reports received which did not meet the farm definition of \$1000 in sales or 1,000 agricultural points were coded as out-of-scope by NASS reviewers.

There were 44 cases for which both the JAS report and its matching census report seemed to contain valid reported data. These were reviewed by FO staff prior to any reinterviews, since it was not considered initially appropriate to contact respondents in cases where NASS staff had simply classified their reports differently. The review revealed that for most of these cases the census records were correctly classified as out-of-scope by the reviewers, and the corresponding June area records were correctly coded as in-scope at the time of the JAS. The FO staff determined that there was a change in the status of the operation between June and the time of the census. Sixty percent of these records showed a slight shift in inventory as the source for the change in operation status. Most of the shift occurred for grain, cattle, and equine operations. In these cases, FO staff recommended a reinterview of the operations to obtain additional important information.

Cases where two or more census reports matched a JAS report were reviewed by FO staff due to confidentiality concerns. When multiple respondents' reports matched a JAS report, NASS was not able to present one respondent with another respondent's report without proper consent. Field Office staff evaluated these cases and provided their best assessment of the situation. There were 536 link groups identified with acreage differences greater than 25 percent, which included two or more census records. The first requirement in this evaluation was to identify any census reports "erroneously" linked to the JAS report. Surprisingly, after removing erroneously linked records, 71 percent of the link groups didn't require any action (reinterview or FO review). Most of these records were linked because of the relationship with management operations and separate individual operations linking to the owners' respective partnerships. The resolution of the remaining 29 percent, or 154 link groups, indicated that 60 percent would require a reinterview in order to reconcile the discrepancies in reported data.

Usually, there was not enough evidence to determine why information was reported differently. Eighteen percent of the groups noted incorrect partnership reporting as the primary reason for duplication on the census. Multiple partners reporting the same information on multiple reports explained most of the duplication. There were a few reports where landlords reported their land incorrectly. Although partnership and landlord issues point to some duplication on the census, a reinterview is still needed in most of these cases to resolve the reporting differences. Further investigation of how partnerships report on the census would help to determine whether this is a significant problem that should be addressed further.

4.1b Reinterview Cases with Scoping Discrepancies

Classification conflicts occurred when census in-scope records matched JAS non-agricultural records; or census out-of-scope records matched JAS in-scope records. There were 67 cases (58 and 9 respectively) that had these scoping differences and were targeted for reinterview.

During the 2002 Classification Error Study, the JAS was assumed as truth. However, operations are misclassified in the census only when their census classification is incorrect. If they are classified differently on the census and JAS, but the JAS is incorrect, then they are not misclassified on the census. Thus, an important aspect of the 2007 CES reinterviews was to ask respondents which source was correct. Respondents indicated whether the census, JAS, neither,

or both reports were correct. The results showed that the census report was correct more often than the JAS report (see Table 5), refuting the assumption used in 2002 that the JAS represented truth. This implies that census misclassification estimates calculated in the 2002 and 1997 CES were most likely significantly overstated, since many of the cases with discrepancies between the census and the JAS were most likely errors on the JAS, NOT misclassification on the census. Of the 67 cases of discrepancies for this study that would have been counted as census misclassifications with the “JAS as truth” assumption, only 15 percent were truly cases of misclassification.

Table 5: Scoping Differences -- Which Source is Correct?

Which Source is Correct	Number (n=67)	Percent
Census is correct	40	59.7%
JAS is correct	10	15.0%
Both are correct	9	13.4%
Neither is correct	8	11.9%

After identifying which source was correct, respondents were asked to provide the reason(s) for the discrepancies. A pre-determined set of reasons was outlined in the survey instrument (Question 5). Additional space was also provided for comments on situations for which the correct reason was not listed. Comments provided by respondents were evaluated to determine if they fit any of the pre-listed reasons. Additional categories were created to group respondents’ comments whenever necessary.

For the cases where the census response was correct, the main reason for the discrepancy was a failure to report agricultural land outside the segment. As a result of this, the operation was made out-of-scope by NASS at the time of the JAS. There were 16 cases in which JAS segments had been improperly screened, resulting in the survey missing valid farm operations. Some examples of the comments validating this problem were “enumerator only observed tract and coded it non-agricultural” and “996 acres were in my segment in June Ag, all were non-agricultural, respondent was not contacted in June.”

Another very important source of error in the JAS was attributed to different types of respondents completing the reports (15 cases). At the time of the CES reinterviews, respondents indicated that a report was answered by a person other than the primary operator. “Ted’s mother did the JAS and did not know the correct answer” and “Wife or help responded” were some of the comments provided by respondents. There were seven reports of respondents that just estimated the acres on the JAS. Finally, there were 10 reports where the discrepancies were due to exclusion of specific types of land, mainly CRP and rented acres. Comments such as “He didn’t consider CRP in June as crop acres” and “Missed reporting CRP in June” indicated that respondents had a difficult time knowing exactly how to report CRP and rented land.

In the cases where the JAS was correct, comments such as “They thought they were not a farm

as they only grow hay” and “Own/Operate 26 acres, keep horses...No longer board horses so we do not consider ourselves farmers” revealed that respondents were incorrectly screening themselves out of the census. There were three reports of respondents who did not consider their operations as farms at the time of the census and as a result did not proceed to complete the census questionnaire as instructed.

There were only two incidences of discrepancies between the JAS and the census that were indicative of true change in an operation, where land was either purchased or sold. Whenever land is purchased it constitutes a true change in an operation, especially if the operator did not own any land at the time of the JAS. Operators that purchased land reported it correctly at the time of the census. However, operators who sold their land after the JAS interview should have reported for the partial year and should have included ALL their land in the census questionnaire. Although selling their land constituted a true change in their operation, it was reported incorrectly. The other cases with discrepancies between JAS and the census were actual errors attributed to the separate operations owned by the respondents (four reports) or operations which were classified as out-of-scope by NASS (e.g., only had woods – no agricultural activity) (two reports).

There were also eight cases where neither the census nor the JAS was correct. The key reasons for the discrepancies were that respondents just estimated the acreage on both reports or they had difficulty reporting rented land.

In summary, in the evaluation of respondents with scoping differences, the census was correct more often than the JAS. This finding refuted the JAS assumption of truth used in 2002. The results showed that a minuscule number of the cases constituted real changes between the census and the JAS. In addition, the amount of misclassification in the census is small and the net effect of misclassification on estimates is even smaller. It is much less than would have been estimated using the JAS assumption of truth. The primary reasons for explained differences in acreage were incorrect screening of tract operations in the JAS, respondent errors in estimation of acreage, the use of proxy respondents, and the exclusion of specific types of land (i.e., CRP, woods, rented).

4.2 Evaluation of Reporting Errors

4.2a Reinterview Cases with Acreage Discrepancies

The majority of discrepancies were acreage differences between the JAS and the census. For these cases, there is no misclassification error, since the operations are classified as farms in both the census and JAS. However, operations should have reported the same number of acres in both if their operation was unchanged in 2007. There were 147 cases with acreage discrepancies. Respondents were first asked whether they had correctly reported their acres on the operation for the JAS, the census, both, or neither (see Table 6). If acreages were correctly reported on both surveys, they were asked why the acreages were different between the two surveys. If the acreages were incorrect on either or both surveys, they were asked why. In both cases, they were given pre-listed reasons to choose from. Multiple responses to questions were allowed. They

were also encouraged to explain their reason in comments.

The evaluation of these cases also showed that the census report was correct more often than the JAS report (46.9 percent vs. 21.8 percent), indicating that the majority of the errors were found on the JAS reports instead of the census report. The results also showed that over 20 percent (31 reports) of the respondents indicated that neither source was correct, while 10.2 percent (15 reports) reported both census and JAS as correct.

Table 6: Acreage Differences -- Which Source is Correct?

Which Source is Correct	Number (n=147)	Percent
Census is correct	69	46.9%
JAS is correct	32	21.8%
Both are correct	15	10.2%
Neither is correct	31	21.1%

Of the 69 responses where the census report was correct, the majority (35 reports) of the respondents said they just estimated their acreage in June as they were *too busy* at the time of the interview. Respondents to the JAS acknowledged estimating their acreage at the time of the survey. They indicated that this was a busy time for them and felt pressured by the interviewers' visit. With the census, they were able to complete the form at their convenience, whereas on the June survey, they felt obligated to respond when the enumerator was there. This was a very surprising fact since personal interviews are expected to yield better results than mail questionnaires. Thus, most of the explained discrepancies where the census was correct and the JAS was in error were due to respondent error. Some of the comments provided were: "The JAS was done outside and I guessed at the acreage" and "I was caught at a busy time, so I gave an approximate answer". The second most common reason for discrepancies in the JAS was that different respondents reported on the two surveys (25 reports), with a spouse or someone less knowledgeable about the operation reporting during the JAS interview. There were comments such as "Danny did the JAS and I did the census." The third most common reason involved the exclusion of specific types of land on the JAS. Twenty June respondents acknowledged leaving off rented, pasture, woods and waste lands. They commented that "They forgot the idle land...rocks and roads" and "The pasture was too low so I did not include it."

There were 32 respondents who reported the JAS was correct. The vast majority of the reasons for discrepancies in these cases also reflected respondent error as a result of estimating acreages. Fifteen respondents said that they estimated acres on the census. Ten respondents pointed out that they neglected to include rented land, and some also excluded waste land or woods. There were a couple of misreports of double cropped and partnership acres, although this was uncommon. Additionally, the effect of having different respondents filtered through as a contributor to errors on the census reports (8 responses). Some of the comments provided were: "The operator added rather than subtracted the acres leased to others" or "Did not include land that he rents". Another comment was "Operator did JAS and part-time helper did the census."

For the 15 cases where both the census and the JAS were reported correctly, an overwhelming majority of the reports (11 cases) constituted true changes in the status of the operations either by the sale, purchase, lease or renting of land. There were seven reports of leased or rented land, three reports of land being purchased, and, one report of land being sold. The respondents commented: “Was renting land, but lost the lease” and “Leased out land after the JAS” or “Bought additional land”.

There were 31 instances where neither report (JAS or census) was correct. The primary reason for both reports being incorrect was that respondents estimated acreages (23 responses). There were also seven cases where certain types of land (i.e., rented, homestead) were excluded on both reports.

In summary, for the group of respondents with acreage differences, the census was correct more often than the JAS, and only a small number of the cases constituted real changes between the census and the JAS. The primary reasons for explained differences in acreage were respondent errors, proxy respondents and the exclusion of specific types of land, such as woods, waste, and CRP.

4.3 Total Acres and Total Land Questions -- Cognitive Research Results

The second part of the 2007 CES was to conduct cognitive research on questions in preparation for the 2012 Census of Agriculture. For the 2002 Census, the total acres question (K46) was edited 31 percent of the time and the total acres of land question (K798) was edited 34 percent of the time. The acreages reported should have been the same for both questions. In addition, the 2005 Census Content Test showed that 32 percent of the time these two fields were not reported the same. Questions were designed on the 2007 CES to address known acreage reporting errors in the Acreage and Land Sections of the 2007 Census questionnaire. These questions were asked of respondents who did not report the same figure for the two acreage questions.

In the 2007 CES reinterview, respondents were first asked what source they used to report their acreage in the Acreage Section of the 2007 Census questionnaire. They were allowed to check multiple answers to this question. There were a total of 214 responses and an overwhelming 49.1 percent of the respondents said they provided their acreage from memory. They checked ‘I know my acreage’ as their response. This was far more than the use of tax records (8.9 percent), FSA records (13.1 percent), operation books (20.6 percent), or any other available source such as deeds or GPS systems (5.6 percent). Table 7 shows a summary of which source was used to report acreage both overall and by group. Respondents were allowed to select multiple answers, so the Table 7 totals do not add to 100 percent.

Estimation of acres from memory was far greater than any of the other possible sources that could have been used to report acreage information. This finding was also true for the 67 cases with scoping differences and the 147 cases with acreage differences.

Table 7: Source Used to Report Acres on Census Questionnaire

Source Used to Report Acres ⁴	Scoping Differences (n=67) Percent	Acreage Differences (n=147) Percent	Total (n=214) Percent
I know my acreage	50.8%	48.3%	49.1%
Tax records	10.5%	8.2%	8.9%
FSA records	6.0%	16.3%	13.1%
Operation books	14.9%	23.1%	20.6%
Other records (i.e., deeds, GPS #s)	1.5%	7.5%	5.6%

On the 2007 Census of Agriculture questionnaire, the total acreage values for codes K46 (total acres of land) and K798 (total acres of land acquired by summing acres by land use) should have been equal. If they were not, respondents were asked why they were different. Table 8 shows the distribution of responses provided. It also provides the breakdown in counts between records with scoping and acreage differences. Respondents were allowed to select multiple answers, so the Table 8 totals do not add to 100 percent.

There were a total of 64 responses, of which 17.2 percent indicated that they did not include the entire operation in K798. About 11 percent of the respondents were not able to explain why the acreages were different. Meanwhile, 9.2 percent of them just overlooked the Land section entirely. From these results, it can be concluded that the majority of the issues were related to the Land Section of the questionnaire (where total acreage was acquired by summing acres by land use – item K798). Other than some respondents excluding rented land in the Acreage Section or their inability to explain the differences between the two sections, the vast majority of the concerns reported dealt directly with the Land Section.

⁴ Respondents were allowed to select multiple answers.

Table 8: K46 Not Equal To K798

K46 should be equal to K798. Why are the acreages different?	Scoping Differences (n=19) Percent	Acreage Differences (n=45) Percent	Total (n=64) Percent
Did not include entire operation in K46	--	11.1%	7.8%
Did not include entire operation in K798	21.1%	15.6%	17.2%
Left out rented land	5.3%	8.9%	7.8%
Reported same acres twice in Land Section	5.3%	4.4%	4.7%
Land Section too complicated, I skipped it	5.3%	2.2%	3.1%
Just overlooked the Land Section	15.8%	6.7%	9.2%
Do not know total acres in my operation	--	--	--
Do not know why they are different	21.1%	6.7%	10.9%

Finally, CES respondents were asked which items in the Land Section were more difficult to report than others. They could check one or more items. Table 8 shows the distribution of responses showing which type of land was most difficult for respondents to report.

There were 214 responses to this question. Of these, 25.7 percent reported woodland not pastured was the most difficult type of land to report. Other land (i.e., waste, pivot edges, CRP) was also problematic for respondents (22 percent). Respondents also noted that ‘cropland idle or used for cover crops or soil improvement but not harvested or pastured’ was a rather intricate description which they found very awkward to report (17.8 percent). Respondents were also asked to comment on why they found it hard to report certain types of land. Some of the comments were: “Easy to misunderstand reporting idle land such as CRP. More difficult to report items such as woods, waste, etc.” Another comment was: “Doesn’t say CRP, doesn’t say your house”. One respondent said: “I thought that section was for only acres where crops were harvested.”

Table 9: Which Type of Land is More Difficult to Report?

Which type of land is more difficult to report?	Scoping Differences (n=67) Percent	Acreage Differences (n=147) Percent	Total (n=214) Percent
Cropland harvested	6.0%	9.5%	8.4%
Cropland where crops failed or abandoned	1.5%	10.2%	7.5%
Cropland in cultivated summer fallow	1.5%	6.8%	5.1%
Cropland idle or used for cover crops or soil improvement but not harvested or pastured	11.9%	20.4%	17.8%
Permanent pasture and rangeland	9.0%	10.9%	10.3%
Woodland pastured	10.5%	10.2%	10.3%
Cropland used only for pasture or grazing	6.0%	10.2%	8.9%
Woodland not pastured	31.3%	23.1%	25.7%
All other land (i.e., waste, pivot edges, CRP)	13.4%	25.6%	22.0%

4. RECOMMENDATIONS

- 1.Improve data collection and classification of operations in the June Agricultural Survey.** Efforts to measure the amount of misclassification in the June Agricultural Survey are already underway. In addition to that research, the use of proxy respondents should be minimized and enumerator training on procedures for proper screening of operations should also be emphasized. If procedures are altered, the next Census Content Test should be used to evaluate them. A sample of respondents from the 2010 JAS should be included in the Census Content Test in December 2010. The data from the two reports should be compared as has been done in the 2007 CES.
- 2.Do not use future Classification Error Surveys to estimate misclassification.** The amount of misclassification in the census is small, especially in terms of net effect on estimates, and the 2007 CES showed that it is much less than would have been estimated using the JAS assumption of truth. More can be gained by examining the reasons for reporting errors in a Classification Error Survey than from attempting to estimate classification error.
- 3.Consider using information about problems with reporting in the Land Section of the census in any redesign of the census form.** Most of the acreage problems within the census were related to the Land Section. This section should be the focus of future content tests.

References

- Abreu, D. A. (2007). Results from the 2002 Classification Error Study. Research and Development Division. RDD Research Report: RDD-07-03.
- Broadbent, K and Iwig, W. (1999), "Record Linkage at NASS Using Automatch". *FCSM Research Conference*, <http://www.fcsm.gov/99papers/broadbent.pdf>
- Johnson, J.V. (2000). Agricultural Census Classification Error Estimation Using an Area Frame Approach. Data Quality Research Section Unpublished Manuscript. National Agricultural Statistics Service, USDA.
- Hanuschak, G.A.; D. Atkinson; W.C. Iwig and V. Tolomeo (1991). History of Reinterview Studies at NASS. Joint Statistical Meetings.
- McCarthy, J. (2007). A Comparison of the Long and Short Forms Used in the 2005 Census Content Test, NASS Research Report #RDD 07-06. Washington, DC.
- Morton, J.E.; P.A. Mullin and P.P. Biemer (2008). Using Reinterview and Reconciliation Methods to Design and Evaluate Survey Questions. *Survey Research Methods*. Vol . No. 2 pp. 75-82.
- The Council on Food, Agricultural and Resource Economics (C-FARE). (2007). *Improving Information About America's Farms & Ranches: A Review of the Census of Agriculture*. Washington, DC.