The Census of Aquaculture

The target population for the census of aquaculture was composed of all aquaculture farm operations that reported any amount of aquaculture activity on their 2017 Census of Agriculture report form. An effort was made to identify additional aquaculture operations of significance from new sources.

The Census of Agriculture

The purpose of a census is to enumerate all objects with a defined characteristic. For the census of agriculture, that goal is to account for “any place from which $1,000 or more of agricultural products were produced and sold, or normally would have been sold, during the census year.”

NASS builds and improves the list on an ongoing basis by obtaining outside source lists. Sources include State and federal government lists, producer association lists, seed grower lists, pesticide applicator lists, veterinarian lists, marketing association lists, and a variety of other agriculture-related lists. NASS also obtains special commodity lists to address specific list deficiencies.

Staff in NASS regional and field offices routinely contact these potential farms to determine whether they meet the farm definition. For the 2017 Census of Agriculture, NASS made a concerted effort to work with community-based organizations not only to improve list coverage for minorities but also to increase census awareness and participation.

List building activities for developing the 2017 CML started in 2014 by updating list information from respondents to the 2012 Census of Agriculture. Between 2015 and 2017, NASS conducted a series of National Agricultural Classification Surveys (NACS) on approximately 1.6million records, which included nonrespondents from the 2012 census and newly added records from outside list sources. The NACS report forms collected information that was used to determine whether an operation met the farm definition. If the definition was met, the operation was added to the NASS list and subsequently to the CML. Addressees that were nonrespondents to a NACS were also added to the CML and identified with a special status code.

NASS uses itsJune Area Survey (JAS) to quantify the number and types of farms not on the CML. The records in the JAS that are not on the CML are said to be in the Not-on-the-Mail List (NML) domain. If a JAS record in the NML domain is determined to be a farm during the census, it is an NML farm. The NML farms are used to measure coverage associated with the census.

The JAS is based on an area frame, which covers all land in the U.S. and includes all farms. The land in the U.S. is stratified by characteristics of the land. A probability sample of segments is drawn within each stratum for the JAS. Segments of approximately equal size are delineated within each stratum and designated on aerial photographs. The JAS sample of segments is allocated to strata to provide accurate measures of acres planted to widely grown crops, farm numbers, and inventories of cattle. Sampled segments in the JAS are personally enumerated. Each operation identified within a segment boundary is known as a tract.

The 2017 JAS sample was increased to improve the farm counts for operations that produced specialty commodities or had socially disadvantaged or minority producers. The total JAS sample consisted of 13,972 segments of which 3,012 were additional segments. This set of additional segmentsis referred to as the Agricultural Coverage Evaluation Survey (ACES)segments. The ACES segments were selected using a multivariate sampling design that targeted specific items at the U.S. level.

During the JAS/ACES enumeration process, each tract is identified as either agricultural or non-agricultural. Each JAS/ACES agricultural tract is identified as a farm or non-farm in June based on the farm definition of $1,000 of sales or potential sales of agricultural products. Non-agricultural tracts are further classified into categories: with farm potential, with unknown farm potential, or with no farm potential. The names and addresses collected in the 2017 JAS/ACES were matched to the CML. Those from the 2017 JAS/ACES that did not match were determined to be in the NML domain and sent a yellow census report form so that they could be differentiated from the green report form sent to those addressees on the CML. Instructions on the census report form directed any respondent who received duplicate forms to complete the CML form and to mail all duplicate forms back together. Those who returned a CML and an NML form had been misclassified as NML and were removed from the NML domain.

The farm/nonfarm status of each NML domain operation was determined based on the reported data in the census form. An operation in the NML domain that was determined to be a farm is referred to as an NML farm. Characteristics of NML farms and their producers provided a measure of the undercoverage of farms on the CML. The percentage of farms not represented on the CML varied by State. In general, NML farms tended to be small in acreage, production, and sales of agricultural products. Farm operations were missing from the CML for various reasons, including the possibility that the operation started after development of the CML, the operation was so small that it did not appear in any agriculture-related source list, or the operation was misclassified as a nonfarm prior to census mailout. The CML was used with the NML in a capture-recapture framework to represent all farming operations across all States in the JAS sample.

Data collection was accomplished primarily by mail, Computer-Assisted Self Interview (CASI) on the Internet, and personal enumeration for special classes of records in the census operations.

NASS regional field offices targeted selected groups of census nonrespondents for in-person enumeration.

Must Case Follow-up. Must cases are known large or unique operations, the absence of which could have significantly affected the accuracy of census results.

Because of the potential importance of Must cases, they were all accounted for and therefore not eligible for nonresponse weighting adjustment.

The NationalNonresponse follow-up activity was designedto focus nonresponse follow-up in a manner that would both reflect the characteristics of the nonresponders and increase response rates. In April 2018, a sample of 249,521 nonrespondents was selected from the remaining 864,260 nonrespondents using a stratified random design. The strata were based on State, county, size of farm, type of farm, producer race, and propensity to respond. . .. Of the 80,504 responses, 51,846 records were identified as being in-scope, resulting in a weighted farm count of 143,847 from the sample.

The NML records were mailed at the same time as the census mailing and received the same follow-up procedures as the census mailing through the first follow-up in mid-February 2018. Beginning in March 2018, CATI was used for nonresponse follow-up for NML nonrespondents.

The edit determined the best value to impute for reported responses that were deemed unreasonable and for required responses that were absent. If an item could not be calculated directly from other current responses, the edit determined whether acreage, production, or inventory items had been reported for that farm on a recent NASS crop or livestock survey. For producers who had not changed in five years,demographics such as race and gender were taken from the previous census. Administrative data from the Farm Service Agency were used for a few items, such as Conservation Reserve Program acreage. When deterministic edit logic and previously-reported data sources were unable to provide a current value, data from a reporting farm of similar type, size, and location were considered. In cases where automated imputation was unable to provide a consistent report, the record was referred to an analyst for resolution.

The complex edit ensured the full internal consistency of the record.

Analysts were provided an additional set of tools, in the form of listings and graphs, to review record-level data across farms. These examinations revealed extreme outliers, large and small, or unique data distribution patterns that were possibly a result of reporting, recording, or handling errors. Potential problems were investigated and, when necessary, corrections were made and the record interactively edited again.

When NASS summarizes data from the census of agriculture, each individual report is typically assigned to a single “principal” county. The principal county is the county in which the majority of an operation’s agricultural products are produced, as reported by the producer. For large operations that have significant production in multiple counties, their reports may be broken up into multiple source counties to more accurately summarize the data.

Although much effort was expended making the CML as complete as possible, the CML did not include all U.S. farms, resulting in list undercoverage. Some farm producers who were on the CML did not respond to the census, despite numerous attempts to contact them. In addition, although each operation was classified as a farm or a nonfarm based on the responses to the census report form, some were misclassified; that is, some nonfarms were classified as farms and some farms were classified as nonfarms. NASS’s goal was to produce agricultural census totals for publication at the county level that were fully adjusted for list undercoverage, nonresponse, and misclassification.

In 2012 NASS used capture-recapture methodology to adjust for undercoverage, nonresponse, and misclassification. This same methodology was implemented for the 2017 Census of Agriculture. To implement capture-recapture methods, two independent surveys were required. The 2017 Census of Agriculture (based on the CML) and the 2017 JAS (based on the area frame) were those two surveys.

A second assumption was that the proportion of JAS farms with a given set of characteristics captured by the census was equal to the proportion of U.S. farms with those same characteristics captured by the census.

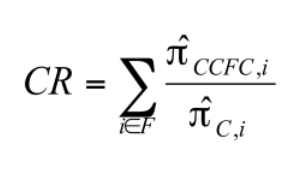
For a farm to be identified as a farm, and thus captured by the census, it must be on the CML, respond to the census report form and, based on the census response, be classified as a farm. Only those nonrespondents included in the nonresponse sample had an opportunity to be captured and had a probability π\_S of being included in the sample; respondents prior to drawing the nonresponse sample had π\_S = 1. Thus, the capture probability π\_C is of interest:

π\_C= π(CML, Responded, Farm on Census|Farm) π\_S

Two types of classification error can occur. First, a farm can be misclassified as a nonfarm. This type of misclassification is accounted for in determining the probability of capture πC. The second type of classification error results when a response to the census is classified as a farm operation when it does not meet the definition of a farm. That is, some farms on the CML may be misclassified from their census report response and may be nonfarms. To account for the misclassification of nonfarms as farms, the probability of a farm on the census being classified correctly must be estimated; that is,

π\_CCFC= π(Farm | Farm on Census)

where CCFC represents Correct Census Farm Classification. To adjust for undercoverage, nonresponse, and misclassification, each CML record classified as a farm based on its response to the census report form was given a weight of the ratio of the estimated probability of correct classification of a farm on the census and the estimated probability of capture ( pi-hat\_CCFC / pi\_hat\_C where the hat symbol (^) denotes an estimate). To estimate the number of farms with a given set of characteristics, the weights of CML records responding as farms on the census and having that set of characteristics were summed. This estimator is referred to as the capture-recapture estimator (CR):



where Fis the set of all CML records classified as farms based on their responses to the census report form.

To estimate the capture and correct census farm classification probabilities, a matched dataset consisting of JAS records and census records was created. Records in the 2017 JAS sample were matched to the 2017 census using probabilistic record linkage. The CML records that matched with JAS tracts represent the Census Sample.

Note: The Census Sample is a subset of the CML records and includes only those records matching a JAS tract. Both agricultural and non-agricultural tracts were included in the matched dataset.

An operation identified as a farm is referred to as in-scope; an operation identified as a nonfarm is referred to as out-of-scope. The farm status based on census responses to either the CML or NML census data collection and the JAS agreed in most cases. In other cases, the JAS and the census (either the CML or the NML) disagreed on whether a record was in-scope or out-of-scope. Such records are said to have conflicting or unresolved farm status. The records with conflicting farm status were sent to NASS regional field offices for review. In each case, efforts were made to determine whether (1) the status had changed between June and December when the census was conducted, (2) the JAS farm status was correct, (3) the census farm status was correct, (4) the records were incorrectly matched, or (5) the farm status could not be resolved.

In 2017, 8.1 percent of the records in the Census Sample had unresolved farm status.

The probability an operation is a farm was estimated for the records with unresolved farm status. Using the 2017 matched dataset, a logistic model of the probability an operation is a farm based on the records with resolved farm status was developed; that is, the operations where the farm (or nonfarm) status agreed between the JAS and the census were used to develop a missing data model, which was then used to resolve farm status. The final missing data model was used to impute the probability that each of the agricultural operations with unresolved farm status is a farm. For the resolved farms and non-farms, the probability of the operation being a farm was 1 and 0, respectively. Five-fold cross-validation was used to develop and to compare competing models. The accuracy of the model was thereby not overstated due to fitting and evaluating the model on the same set of data. To ensure that each of the cross-validation samples covered the U.S., the five cross-validation samples of JAS segments were drawn within State-stratum combinations. Characteristics of the JAS tracts were considered as potential covariates in the model. Because limited information is available for JAS nonfarm tracts, other covariates considered included county-level socio-demographic variables from the most recent U.S. population census, segment-level data from the Cropland Data Layer, the county-level rural-urban code, state-level response rates, an indicator for records that are thought to be out-of-business, and an indicator for records in the national non-response sample. The sample weight associated with each JAS tract was multiplied by the probability of being a farm. This adjusted weight was used in all subsequent modeling.

For a farm to be identified as a farm, and thus captured, by the census, it must be on the CML, respond to the census report form and, based on the census response, be classified as a farm. These adjustments are dependent. Further, those non-respondents at the time the nonresponse sample was drawn had a known probability π\_S of being included in the sample; respondents before the sample was drawn had π\_S = 1. Therefore, the probability of capture *π\_C* may be written as

π\_C = π(CML, Responded, Farm on Census|Farm) π\_S

π\_C = π(CML|Farm)π(Responded|CML, Farm)π(Farm on Census|CML, Responded, Farm) π\_S

The probability of being included in the sample π\_S is known for all responding farms. The other terms in the probability of capturing a farm depend on the characteristics of the farm. Using five-fold cross-validation, three logistic models were developed based on the matched dataset. The first model estimated the probability of a farm being on the CML. The second model estimated the probability that a farm on the CML responded to the census report form. The final model estimated the probability that a farm that was on the CML and responded to the census was identified as a farm based on its response. The probability that a farm is captured by the census of agriculture is then the product of the three conditional probabilities that a farm is on the CML, responds, and is identified as a farm.

Note 1: Responses were required for Must cases. These operations were only excluded in modeling the probability of a farm responding given that it was on the CML.

An operation is misclassified if: (1) it meets the definition of a farm, but is classified as a nonfarm on the census or (2) it does not meet the definition of a farm, but is classified as a farm on the census. The first type of misclassification is accounted for when modeling the probability of capture. An adjustment is still needed for the misclassification of nonfarms as farms. As with farm status and capture, the probability of this misclassification depends on an operation’s characteristics. Thus, a final logistic model was developed. Given that an operation was classified as a farm on the CML, the probability of its being a farm was modeled based on its characteristics. Five-fold cross-validation was used to ensure that the model was not over-fitted.

Each operation identified as being in-scope on the CML was given a weight equal to the probability of misclassification divided by the probability of capture. This weight accounted for undercoverage, nonresponse, both types of misclassification, and the nonresponse sample.

The record weighting processes were initially applied at the State level to produce adjusted estimates of farm numbers and land in farms for 63 different categories of 8 characteristics of the farm operation or the farm producer -- value of agricultural sales (9); age (2); female; race (3); Hispanic origin of principal farm producer; 4 sales categories for each of 10 major commodities (40); and farm type groups (7). The State-level number of farms and land in farms were two additional adjusted estimates, resulting in 65 categories. To reduce the intercensal variation at the

State level, the State targets were smoothed by averaging the 2017 estimates from capture-recapture and the published 2012 State estimates with the restrictions that the smoothed targets were within two standard errors of the capture-recapture estimates. The smoothed State targets were rescaled so that they summed to the national capture-recapture estimates.

These State estimates were general purpose in that they did not provide any control over expected levels of commodity production of the individual farm operation. As a result of this limitation, the procedures could have over-adjusted or under-adjusted for commodity production. To address this, a second set of variables, known as commodity targets, was added to the calibration algorithm. These targets were commodity totals from administrative sources or from

NASS surveys of nonfarm populations (e.g. USDA Farm Service Agency program data, Agricultural Marketing Service market orders, livestock slaughter data, cotton ginning data). The introduction of these commodity coverage targets strengthened the overall adjustment procedure by ensuring that major commodity totals remained within reasonable bounds of established benchmarks.

Each State was calibrated separately. The calibration algorithm addressed commodity coverage. The algorithm was controlled by the 65 State farm operation coverage targets and the State commodity coverage targets. Because calibration targets are estimates subject to uncertainty, NASS allowed some tolerance in the determination of the adjusted weights.

Rather than forcing the total for each calibration variable computed using the adjusted weights to equal a specific amount, NASS allowed the estimated total to fall within a tolerance range.

Tolerance ranges for the farm operation coverage targets were determined differently from the commodity targets. The tolerance range for the 65 State farm operation coverage targets was the estimated smoothed State total for the variable plus or minus one standard error of the capture-recapture estimate. This choice limited the cumulative deviation from the estimated total for a variable when State totals were summed to a U.S. total. Commodity coverage targets with acceptable ranges were established based on the administrative source for each State. Ranges were not necessarily symmetric around the target value.

To ensure that all subdomains for which NASS publishes summed to their grand total, integer weights were produced by a discrete calibration algorithm. This eliminated the need for rounding individual cell values and ensured that marginal totals always added correctly to the grand total. If a weight was initially not in the interval [1,6], it was trimmed so that in was in that interval. That is, adjusted weights less than 1 were set to 1, and those greater than 6 were set to 6. The remaining non-integer weights were then rounded sequentially to reduce the distance of the estimated totals from the targets.

Calibration adjustments began with the computation of a priority index for each record. The priority index was the absolute value of the gradient of the relative error associated with increasing or decreasing a record’s weight by one. The record with the highest priority index was then selected as a candidate to increase or decrease its weight by one to reduce the cumulative distance from the targets as measured by the relative error. If the new value produced an improvement and satisfied the range restrictions, the weight was updated and new priorities were assigned; otherwise, the record with the next highest priority index was processed. This process was iteratively performed until convergence was attained. Because census data collection was assumed to be complete for very large and unique farms, their weights were controlled to 1 during the calibration adjustment process. For all other farms, the final census record weights were forced to be an integer number in the interval [1, 6]. The calibration process considered all targets simultaneously through the priority index. Although calibration was seldom able to adjust weights so that all State targets were met, all targets were brought collectively as close to the targets as possible.

The proportions of selected census data items that were due to coverage, response, and classification adjustments are displayed in Tables A and C.

Table C. Summary of Coverage, Nonresponse, and Misclassification Adjustments for SC: 2017

Farms Land in farms Sales

Total (number) 24,791 4,744,913 acres $3,008,739,000

Standard error 1,346 347,006 acres $139,270,000

Adjustment as percent of total 44.8% 28.6% 12.0%

% adjustment from coverage 16.6% 8.5% 2.6%

% adjustment from nonresponse 17.6% 13.3% 7.2%

% adjustment from misclassification 10.6% 6.8% 2.2%

**Non-disclosure of sensitive information**

any total that would reveal an individual’s information or allow it to be closely estimated by the public. Farm counts are not considered sensitive and are not subject to disclosure controls. Cell suppression was used to protect the cells that were determined to be sensitive to a disclosure of information.

Based on agency standards, data cells were determined to be sensitive to a disclosure of information if they failed either of two rules. The threshold rule failed if the data cell contained less than three operations. For example, if only one farmer produced turkeys in a county, NASS could not publish the county total for turkey inventory without disclosing that individual’s information. The dominance rule failed if the distribution of the data within the cell allowed a data user to estimate any respondent’s data too closely. For example, if there are many farmers producing turkeys in a county and some of them were large enough to dominate the cell total, NASS could not publish the county total for turkey inventory without risking disclosing an individual respondent’s data. In both of these situations, the data were suppressed and a “(D)” was placed in the cell in the census publication table. These data cells are referred to as primary suppressions.

Since most items were summed to marginal totals, primary suppressions within these summation relationships were protected by ensuring that there were additional suppressions within the linear relationship that provided adequate protection for the primary. A detailed computer routine selected additional data cells for suppression to ensure all primary suppressions were properly protected. These data cells are referred to as complementary suppressions. These cells are not themselves sensitive to a disclosure of information but were suppressed to protect other primary suppressions. A “(D)” was also placed in the cell of the census publication table to indicate a complementary suppression. A data user cannot determine whether a cell with a (D) represents

a primary or a complementary suppression.

Regional field office analysts reviewed all complementary suppressions to ensure no cells had been withheld that were vital to the data users. In instances where complementary suppressions were deemed critically important to a State or county, analysts requested an override and a different complementary cell was chosen.

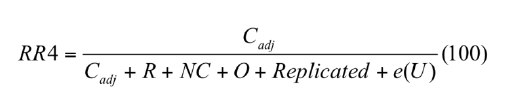
**Census Quality**

It is not likely that either the CML includes all operations that meet the definition of a farm or that all those that do meet the definition of a farm respond to the census inquiry. The goal is to publish data with a high level of quality. The quality of a census may be measured in many ways. One of the first indicators used is a measure of the response to the census data collection as it has generally been thought that a high response rate indicates more complete coverage of the population of interest. This is a valid assumption if the enumeration list, the CML here, has complete coverage of the population of interest. In the case of the census of agriculture, the definition requiring advance knowledge of sales makes achieving a high level of coverage difficult. To ensure that the census of agriculture is as complete as possible, records are included that might not meet the census definition of a farm – in fact, almost 50 percent more records than the anticipated number of qualifying farm operations were included in the 2017 CML. A second indicator of quality then is the coverage of the farm population by the CML. Other indicators of quality relate to the accuracy and completeness of the data, and the validity of the procedures used in processing the data.

**Census Response Rate**

The response rate for the 2017 Census of Agriculture CML was 71.8 percent, as compared with the 2012 Census of Agriculture’s response rate of 74.6 percent and 78.2 percent for the 2007 Census of Agriculture.

The 2017 Census of Agriculture’s response rate used the fourth response rate formula (RR4) from the American Association of Public Opinion Research’s Response Rate Standard Definitions manual:



where

*Cadj* = number of fully and partially completed records, excluding replicated records

*R* = number of explicit refusals

*NC* = number of non-contacted operations known to be eligible

*O* = number of other types of nonrespondents

*Replicated* = number of replicated records

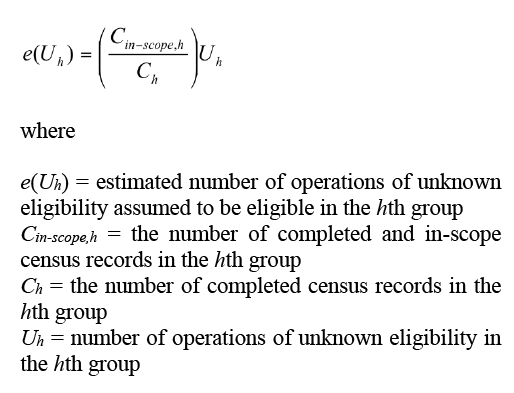
*U* = number of operations of unknown eligibility

*e*(*U*) = estimated number of operations of unknown eligibility assumed to be eligible

Records were classified into the above variables based on the combination of their active status (AS) codes, in-scope status, and replication status. Active status refers to the eligibility status of records for selection on the CML. All replicated records were considered to be a form of nonresponse and were classified into other nonrespondents; in-scope status was considered immaterial.

Certain active status classifications indicated records of unknown agricultural status. These classifications included records to be removed from the CML but had data from outside sources indicating agricultural activity, new records from outside data sources, non-respondents and refusals to the NACS, records for regional office handling only, and records with Farm

Service Agency or Conservation Reserve Program data on operations that are not owned by the principal producer. These records were stratified (grouped) based on their probabilities of being in-scope had they responded. The estimated number of in-scope non-respondents was calculated for the *h*th stratum (group) by the following formula:



**Census Coverage**

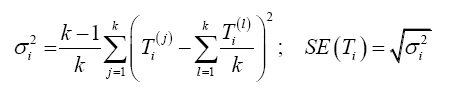
As a side-product of the statistical adjustment used to account for undercoverage, nonresponse of farms on the CML, and misclassification of responses to the census, the proportion of the adjustments due to each of those factors can be derived. The percentages of final census estimates due to adjustments for undercoverage, nonresponse, and misclassification as well as the total percent adjustment for selected items are displayed in Tables A and C.

**Measured Errors in the Census Process**

Although the census of agriculture does not inherently rely on a sample, NASS used a national nonresponse sample as part of its follow-up efforts in 2017. In addition to the uncertainty introduced by the nonresponse sample, NASS uses statistical procedures in compiling the CML, in its data collection procedures, in data editing and processing, and in compiling the final data. Additionally, it uses statistical procedures to both measure errors in the various processes and in making adjustments for those errors in the final data. One example is the statistical process used to account for under-coverage, nonresponse of farms on the CML, and misclassification of responses to the census. The basis of the under-coverage adjustment is the capture-recapture procedure that uses the area sample enumeration from the JAS. The largest contributors to error in the census estimates are due to the adjustments for nonresponse, under-coverage, misclassification, calibration, and integerization.

In conducting the 2017 Census of Agriculture, efforts were initiated to measure error associated with the adjustments for farm operations that were not on the CML, for farm operations that were on the CML but did not respond to the census report form, and for farms and non-farms that were misclassified as non-farms and farms, respectively, for calibration. These error measurements were developed from the standard error of the estimates at the national, State, and county levels and were expressed as coefficients of variation (CVs) at the national and State levels and as generalized coefficients of variation (GCVs) at the county levels.

The standard error of an estimate is an estimate of the standard deviation of the sampling distribution of the estimator. In each case, standard errors were computed using an approach based on a combination of group jackknife and bootstrap methodologies. To conduct the jackknifing, *k* = 10 mutually exclusive and exhaustive groups of JAS segments were formed. The groups were selected using a stratified random design so that each group reflected the survey design, including State and agricultural strata within a State. The weight of record *i* in jackknife group *j* is *CRi^(j )*for *j* = 1, 2, …, *k*. Based on these weights, a group jackknife estimator to estimate the variance would account for the uncertainty associated with modeling the capture-recapture probabilities. To account for the additional uncertainty due to calibration, the weights within each jackknife group were transformed through bootstrap simulation; these transformed weights are called calibration-adjusted-jackknife weights. The full dataset, which is composed of the records of all responding farms on the CML, is calibrated as described in the Calibration section, and the final calibration-adjusted weight of record *i* is denoted by *ŵi*. For each record *i* in jackknife group *k*, the calibration-adjusted-jackknife weights of that record can be approximated as *wi^(j)=ai^(j)CRi^(j)* where *ai^(j) ~N(*1*,( ŵi* – 1) / *ŵi*). The bootstrap process simulated the value of the adjustment *ai^(j)* for each record on the CML to obtain the calibration-adjusted-jackknife weights. For a given data item, such as the number of farms, the estimate *T^(j)* was computed at the specified geographical level, such as nation, State, or county, using the (*k* – 1) groups remaining after deleting the calibration-adjusted jackknife group *j*. Estimates of the variance and standard error associated with the estimator *Ti* are then, respectively,

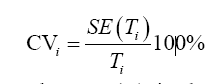


Increasing *k* improves the estimate of the variance but, as *k* increases, the observations become too sparse to reflect the survey design and to provide countrywide coverage. Ten (10) calibration-adjusted jackknife groups were used to provide standard errors for 2017

State and national estimates. For the estimate of the number of farms with a given set of characteristics, only the CML records with those characteristics were used to obtain the overall estimate as well as the estimates from each calibration-adjusted jackknife group.

Note that the calibrated jackknife groups were only constructed once, and different subsets of the records were used to compute estimates and standard errors for the data items.

The CV is a measure of the relative amount of error associated with the sample estimate:



where *SE*(*Ti*) is the standard error of the capture-recapture estimate for data item *i*. This relative measure allows the reliability of a range of estimates to be compared. For example, the standard error is often larger for large population estimates than for small population estimates, but the large population estimates may have a smaller CV, indicating a more reliable estimate. For county-level estimates, a generalized coefficient of variation (GCV) was determined for each estimate within a State. A generalized variance function relates a function of the variance of an estimator to a function of the estimator. Within a State, the standard error of an estimate for a data item was often found to be linearly related to the estimate of that item with an intercept of zero. Based on this modeled relationship, the GCV is the slope of the line relating the standard error to the estimate, multiplied times 100 to represent the GCV as a percentage.

The standard error is the product of the CV (or GCV for county estimates) and the estimate divided by 100. As an example, if the GCV for a State is 25 percent

and a county’s estimate is 4, then the standard error is

25(4)/100 = 1. The standard error of an estimated data

item from the census provides a measure of the error

variation in the value of that estimated data item based

on the possible outcomes of the census collection,

including variants as to who was on the CML, who

returned a census form, who was misclassified either

as a farm or as a nonfarm, and the uncertainty

associated with calibration and integerization. With

95 percent confidence, an estimate is within two

standard errors of the true value being estimated. For

this example, with 95 percent confidence, the estimate

of 4 is within 2(1) = 2 of the true county value.

Table B presents the fully adjusted estimates with the

coefficient of variation for selected items.

As noted in the previous section, sampling errors can

be introduced from the coverage, nonresponse and

misclassification adjustment procedures. This error is

measureable. However, nonsampling errors are

imbedded in the census process that cannot be directly

measured as part of the design of the census but must

be contained to ensure an accurate count. Extensive

efforts were made to compile a complete and accurate

mail list for the census, to elicit response to the

census, to design an understandable report form with

clear instructions, to minimize processing errors

through the use of quality control measures, to reduce

matching error associated with the capture-recapture

estimation process, and to minimize error associated

with identification of a respondent as a farm operation

(referred to as classification error). The weight

adjustment and tabulation processes recognize the

presence of nonsampling errors; however, it is

assumed that these errors are small and that, in total,

the net effect is zero. In other words, the positive

errors cancel the negative errors.

Incorrect or incomplete responses to the census report

form or to the questions posed by an enumerator can

introduce error into the census data. Steps were taken

in the design and execution of the census of

agriculture to reduce errors from respondent

reporting. Poor instructions and ambiguous

definitions lead to misreporting. Respondents may not

remember accurately, may estimate responses, or may

record an item in the wrong cell. To reduce reporting

and recording errors, the report form was tested prior

to the census using industry accepted cognitive testing

procedures. Detailed instructions for completing the report form were provided to each respondent.

Questions were phrased as clearly as possible based

on previous tests of the report form. Computerassisted

telephone interviewing software included

immediate integrity checks of recorded responses so

suspect data could be verified or corrected. In

addition, each respondent’s answers were checked for

completeness and consistency by the complex edit

and imputation system.

Processing of each census report form was another

potential source of nonsampling error. All mail

returns that included multiple reports, respondent

remarks, or that were marked out of business and

report forms with no reported data were sent to an

analyst for verification and appropriate action.

Integrity checks were performed by the imaging

system and data transfer functions. Standard quality

control procedures were in place that required that

randomly selected batches of data keyed from image

be re-entered by a different operator to verify the work

and evaluate key entry operators. All systems and

programs were thoroughly tested before going on-line

and were monitored throughout the processing period.

Developing accurate processing methods is

complicated by the complex structure of agriculture.

Among the complexities are the many places to be

included, the variety of arrangements under which

farms are operated, the continuing changes in the

relationship of producers to the farm operated, the

expiration of leases and the initiation or renewal of

leases, the problem of obtaining a complete list of

agriculture operations, the difficulty of contacting and

identifying some types of contractor/contractee

relationships, the producer’s absence from the farm

during the data collection period, and the producer’s

opinion that part or all of the operation does not

qualify and should not be included in the census.

During data collection and processing of the census,

all operations underwent a number of quality control

checks to ensure results were as accurate as possible.

All item nonresponse actions provide another

opportunity to introduce measurement errors.

Regardless of whether it was previously reported data,

administrative data, the nearest neighbor algorithm, the fully conditional specification method, or

manually imputed by an analyst, some risk exists that

the imputed value does not equal the actual value.

Previously reported and administrative data were used

only when they related to the census reference period.

A new nearest neighbor was randomly selected for

each incident to eliminate the chance of a consistent

bias.

The process of building and expanding the CML

involves finding new list sources and checking for

names not on the list. An automated processing

system compared each new name to the existing CML

names and “linked” like records for the purpose of

preventing duplication. New names with strong links

to a CML name were discarded and those with no

links were added as potential farms. Names with weak

links, possible matches, were reviewed by staff to

determine whether the new name should be added.

Despite this thorough review, some new names may

have been erroneously added or deleted. Additions

could contribute to duplication (overcoverage)

whereas deletions could contribute to undercoverage.

As a result, some names received more than one

report form, and some farm producers did not receive

a report form. Respondents were instructed to complete one form and return all forms so the

duplication could be removed.

Another chance for error came when comparing June

Area Survey tract producer names to the CML. Area

producers whose names were not found on the CML

were part of the measure of list incompleteness, or

NML. Mistakes in determining overlap status resulted

in overcounts (including a tract whose producer was

on the CML) or undercounts (excluding a tract whose

producer was not on the CML). All tracts determined

to not be on the list were triple checked to eliminate,

or at least minimize, any error. NML tract producers

were mailed a report form printed in a different color.

In order to attempt to identify duplication, all

respondents who received multiple report forms were

instructed to complete the CML version and return all

forms so duplication could be removed.

Records in the 2017 JAS were matched to the 2017

census using probabilistic record linkage. The records

of operations with differing farm status were sent out

to be reviewed by NASS regional field offices. If farm

status could not be resolved, the probability of an

operation being a farm was imputed using a missing

data model. The uncertainty associated with this

estimate, with the exception of model uncertainty,

was accounted for, but errors not found through this

process were not.

The Irrigation and Water Management Survey

**Irrigated farms.** Irrigated farms or ranches are those with any agricultural land irrigated by any artificial or controlled means in the specific calendar year. The acreage irrigated may vary from a very small portion of the total acreage in the farm or ranch to irrigation of all agricultural land in the farm or ranch. This includes reclaimed water and livestock lagoon wastewater distributed by sprinkler or flood systems.

**Irrigated land.** Irrigated land is defined as “all land watered by artificial or controlled means.” No attempt was made to define the degree or intensity of irrigation. Therefore, the data for irrigated land include land with as little as one-half inch of water applied as well as land with several acre-feet of water applied.

**Target population.** The target population includes all farms that irrigated sometime 2013-2017 and 2018, excluding institutional, research, and experimental farms. Operations with irrigation capabilities may not irrigate depending on the amount of rainfall for a particular year or geographic area. New irrigators in 2018 (not included in the 2017 census) were excluded from the survey.

**Sampling.** A certainty stratum, with farms selected with probability one, was included for each State to ensure that the major irrigators in each State were sampled. The remaining strata were sampled systematically by irrigated acreage. The stratification boundaries varied among the States and were dependent on the distribution of total acres irrigated within the State.

**Data collection.** Data were collected by mail, Computer-Assisted Web Interviewing (CAWI) via the Internet, telephone enumeration, and personal enumeration. Enumeration methods used in the 2018 survey were similar to those used in the 2013 survey. The report form was mailed to all the producers in the sample that reported irrigation in the 2017 Census of Agriculture. The initial mail packets were sent February 2019, and included a labeled report form, an instruction booklet, an instruction letter, and a return envelope. Mail-out packet preparation, initial mail-out, and one follow-up mailing to non-respondents. Telephone follow-up began April 2019 to non-respondents. Agency contacts in field offices collected data for operations which were scheduled for contact for other agricultural surveys.

**Estimation.** The estimation methodology consisted of two weighting components that made up the total survey weight. The first component was the fully adjusted weight pulled in from the 2017 Census of Agriculture. This weight accounted for any list incompleteness and under-coverage from the 2017 census. The second component was the sampling rate used for the 2018 Irrigation and Water Management Survey. This expansion factor was the inverse of the selection probability for the sample farms in a stratum, reweighted at the stratum level to account for whole-farm nonresponse. The nonresponse adjustment factor used to reweight the expansion factor was the ratio of the number of sample farms in a stratum to the number of sample farms that responded to the survey in that stratum. The assumption underlying this weighting approach to survey nonresponse was that survey respondents and non-respondents within a stratum constitute a homogeneous population, thus allowing respondents to represent non-respondents.

Any tabulated item that identifies data reported by a respondent or allows a respondent’s data to be accurately estimated or derived, was suppressed and coded with a ‘D’. However, the number of farms reporting an item is not considered confidential information and is provided even though other information is withheld.

The accuracy of data products may be evaluated through sampling and non-sampling error. The measurement of error due to sampling in the current period is evaluated by the coefficient of variation (CV) for each estimated item. Non-sampling error is evaluated by response rates and the percent of the estimate from respondents.

Coefficient of variation is a measure of the relative amount of error associated with a sample estimate. Specifically, it is the standard error of a point estimate divided by that estimate, generally multiplied times 100 so that it can be reported as a percentage. This relative measure allows the reliability of a range of estimates to be compared. For example, the standard error is often larger for large population estimates than for small population estimates, but the large population estimates may have a smaller CV, indicating a more reliable estimate. Every estimate for the 2018 Irrigation and Water Management Survey has a corresponding CV published with it.

Table C. Coefficient of Variation for General Irrigation Data: 2018 (South Carolina)

Irrigated farms: 6.5%

Land in farms: 18.5%

Acres irrigated (total): 33.6%

Acres irrigated (cropland harvested in the open): 34.3%

Acre-feet applied: 36.7%

Energy expense for pumps: 33.0%

Expenditure expenses: 47.5%

Pumps, all types: 25.7%

Well pumps: 26.3%

Table D. Coefficient of Variation for Horticultural Irrigation Data: 2018 (South Carolina)

Irrigated horticultural operations: 20.1%

Irrigated area, acres in the open: 73.8%

Irrigated area, square feet under protection: 85.9%

**Cropland harvested in the open.** Acres that included harvested field crops and hay along with land used for vegetables, orchards, citrus groves, vineyards, berries, nuts, cultivated Christmas trees, short-rotation woody crops, nursery and other horticultural crops grown in the open. Does not include land used for growing horticultural crops under protection nor cropland reported as failed or used for cover crops.

If two or more crops were harvested from the same land during the year (double cropping), the acres were counted for each crop. Therefore, the total acres of all crops harvested could exceed the acres of cropland harvested. An exception to this procedure was hay crops.

**All other crops.** Data relate to any non-vegetable crops not having a specified code on the 2018 report form. Crops such as grass seed, sunflower seed, sugarcane, etc. were included in other field crops. Horticulture in the open was recorded under “Other cropland” during 2013 while it was recorded as its own commodity during 2018. Therefore, “Other cropland” is not comparable between 2018 and 2013.

**Horticultural crops:**

*Cultivated Christmas trees and short rotation woody crops in the open.* Includes all trees to be cut in 2018 or later years. Irrigated live Christmas trees were reported in nursery crops.

*Floriculture and bedding crops.* Includes annual bedding/garden plants, herbaceous perennials, cut flowers and cut cultivated greens, foliage plants for indoor or patio use, potted flowering plants, and other floriculture type crops.

*Food crops grown under protection.* Includes all food crops that were grown in a greenhouse or under some sort of structure that regulated light, shade, temperature, etc. No food crops grown in open fields were reported for this crop type.

*Mushroom crops under protection.* Includes all mushroom species that were irrigated in 2018. Logs were converted to and reported in square feet.

*Nursery crops.* Includes deciduous shade and flowering trees, broadleaf and coniferous evergreens, live Christmas trees for sale as potted trees or balled and burlapped, fruit and nut trees and plants, ornamental grasses, palms for landscaping, shrubs, vines, aquatic plants, and other woody ornamentals.

*Propagative materials.* Includes dry bulbs, corms, tubers, and rhizomes; cuttings, seedlings, liners, and plugs; flower seeds; vegetable seeds; vegetable transplants; and tobacco transplants.

*Sod in the open.* Includes all irrigated sod, sprigs, or plugs.

*Other horticultural crops.* This category includes all crops that are primarily considered ornamental or horticultural and are not listed on the report form. Fruit, nut, and vegetable crops grown in the open are not considered horticultural crops.

**Reclaimed water.** Reclaimed water is wastewater that has been treated for non-potable reuse purposes. Sources include municipal, industrial, off-farm livestock operations, and other reclaimed water sources. Water from off-farm livestock facilities, municipal, industrial, and other reclaimed water sources were reported as off-farm supplies. While reclaimed water from on-farm livestock facilities were reported as on-farm surface water.

**Recycled water.** Recycled water is the reuse of surface or ground water that has already been used to irrigate a crop on the operation. Recycled water use was reported as on-farm surface water.

Compare irrigators (table 1), irrigation volume by source (table 4), irrigation by quantity (table 7), and irrigation wells (table 8), with SCDHEC data.