

## Canada – Mexico – USA

### Agricultural Data Concepts and Metadata Group

#### 1. DEFINITIONS

Agricultural statistics are based on various surveys and censuses conducted in Canada, Mexico and the United States. The three countries run their respective agricultural statistics programs through farming information obtained from collection units.

##### 1.1 Canada

The main agricultural units used to collect Canadian agriculture statistics are:

Operation: A Canadian agricultural operation is the lowest operating level in a business which produces agricultural products intended for sale for which an independent income and expense statement is available. Provincial boundaries are respected.

Operator: A Canadian agriculture operator is a person responsible for the day-to-day decisions made in the agricultural operation, potentially with other operators.

##### 1.2 Mexico

The main agricultural units used to collect Mexican agriculture statistics are:

Production Unit: A production unit includes all plots of land that are operated by the same administration. There are two main types of properties: private properties and communal properties known as “*ejidos*”. The communal properties include many subsistence units that have no intention to sell agricultural products.

Producer: The person or persons who manage the production unit, that is to say that this group of people is responsible for all or most decisions for the operation of the farm such as planting and harvesting.

##### 1.3 United States

The main agricultural units used to collect American agriculture statistics are:

Farm: Any place which produce and sold or normally would have produced or sold \$1,000 worth of agricultural products during the year.

Farmer: The person responsible for all or most of the day-to-day decisions such as planting, harvesting, feeding, or marketing for the farm operation. The operator could be the owner, hired manager, cash tenant, share tenant or a partner.

## **2. AGRICULTURE STATISTICS PROGRAMS**

The Agriculture Statistics Programs maintained by the three countries are based on a mixture of surveys, censuses and administrative sources.

### **2.1 Canada**

The Canadian Agriculture Statistics Program collects, compiles, analyzes and publishes a wide range of information on the agricultural sector in Canada and in the individual provinces. The program includes annual, quarterly and seasonal surveys on crops, horticulture, livestock, farm finances, and aquaculture, as well as a Census of Agriculture conducted every five years. In addition, to reduce respondent burden and survey cost wherever possible, administrative data from federal and provincial governments and agencies, marketing boards and producer organizations are used in place of agriculture surveys.

The agriculture surveys vary in size and content. The large regular surveys are (a) the Crops Survey that is conducted six times per year to collect information on seeding intentions, area seeded, yield, production and stocks of grains, (b) the Livestock Survey conducted in January and July every year to collect inventories of cattle, hogs, and sheep as well as other types of livestock, (c) the Hogs Survey collected in April and October to collect more detailed information on the hogs industry, (d) the Atlantic Survey conducted in June and November to collect both crops and livestock information in the Atlantic provinces since they are not covered by the regular crops or livestock surveys, (e) the annual Farm Financial Survey conducted in the spring to collect information about revenues, expenses and properties, as well as physical characteristics, (f) the Fruit and Vegetables Survey conducted in May and November to collect area and production of fifty different fruits and vegetables, and (g) the annual Greenhouse, Sod and Nursery Survey conducted in February to obtain a view of the general economic situation within these industries. Smaller surveys include (h) the Potato Area and Yield Survey, (i) the Potato Prices Survey, (j) the Hay Prices Survey, (k) the Corn and Soybean Survey, (l) the Mushroom Survey, (m) the Honey Survey, (n) the Maple Products Survey, and (o) the Eggs Survey.

In addition, the Farm Update Survey (FUS) is conducted annually, principally to identify births to be added to the Farm Register and to the survey frames. This survey makes use of tax return forms to identify units with a high probability of being new farms. Prior to the 2008 reference year the largest of these potential birth units were selected for the survey, but starting in 2008 the survey will use a probabilistic approach.

Finally, occasional surveys are conducted on a cost recovery basis in response to specific requirements from other government agencies or external clients. Although most are one-time surveys, their scope is rarely negligible. The Farm Environment Management Survey, the Livestock Farm Activity Survey, the Crop Protection Survey, and the Agricultural Water Use Survey are examples.

The Agriculture Statistics Program is completed by the Tax Data Program, the Farm Income Statistics Program, Spatial Analysis and Geomatic Applications and rural and agriculture research.

## **2.2 Mexico**

The Mexican agricultural statistics program collects, compiles, analyzes and publishes a wide range of information on the agricultural sector in Mexico. The program uses administrative registers to obtain most crop and livestock production; seasonal surveys of the most important crops for production, prices and production costs and remote sensing for the estimated sown area of the main crops. In addition, an agricultural census is conducted every ten years. The last census was conducted in 2008.

The administrative information is combined, analyzed and validated by two information systems: the agriculture information (SIACAP) and livestock information (SIPCAP) systems. In particular, the agriculture information includes the sown area, harvested area, damaged area, observed and estimated harvests, the observed and estimated production and the rural average price. The livestock information includes the inventory, weight and reported prices of cattle, pigs, sheep and goats, the weight and price of poultry, production, the weight and price of cattle, pig, sheep, goat and poultry meat and the production and price of agriculture and livestock products such as milk, eggs and others.

The sample surveys take place in the top producing states whose combined production accounts for 80% of the total production. Within these states, they take place in the top producing cultivated areas whose combined production accounts for 80% of the state's total production. Seasonal interviews with the producers and physical measurements in the plots are used to estimate the production of the production units. The main products in Mexico are corn, beans, sorghum, wheat and others. In terms of production value, vegetables are significant.

Lastly, the sown area is estimated by using field information and satellite-based images.

## **2.3 United States**

The USDA's National Agricultural Statistics Service (NASS) conducts hundreds of surveys every year and prepares reports covering virtually every aspect of U.S. agriculture. Production and supplies of food and fiber, prices paid and received by farmers, farm labor and wages, farm finances, chemical use, and changes in the demographics of U.S. producers are only a few examples.

The program includes surveys conducted annually, quarterly, monthly, seasonally and in a few instances weekly. In addition, a Census of Agriculture is conducted every five years. Where feasible, administrative data from federal and state agencies and private sources are used to supplement data collection and reduce respondent burden and survey cost.

### *2.3.1 Crop Statistics Program*

Surveys conducted by NASS vary in size and scope. The crops/stocks surveys are the major source of data for estimates of crop acreage, yields and production, and quantities of grain and oilseeds stored on farms. The surveys are conducted quarterly in March, June, September and December. Farmers planting intentions are collected in March, acres planted and acres expected for harvest are collected in June, small grains acres harvested

and production are collected in September, and row crop and hay production are collected in December. Information on grains and oilseeds stored on the farm for major commodities are collected in all four quarters, while specialty crops are collected once annually in selected producing states.

An Agricultural Yield Survey provides farmer reported survey data of expected crop yields used to forecast and estimate crop production levels throughout the growing season. The survey is conducted monthly from May through November. Small grains data are collected from May through August. Row crop data are collected from August through November. Hay yield data are collected in August and October with hay stocks collected in May. Tobacco data are collected from May through November. Objective Yield measurement surveys are conducted during the same periods for major crops in major producing states. Other surveys varying in frequency include the Floriculture survey, and the Fruit and Nut survey.

### *2.3.2 Livestock Statistics Program*

Cattle, Sheep and Goat inventory surveys are conducted bi-annually in January and July. Cattle on Feed surveys are conducted monthly. Milk production and Dairy Products surveys are conducted monthly. Hog surveys are conducted quarterly in March, June, September and December. A Broiler Hatchery survey of the number of eggs placed in hatcheries for broiler production is conducted weekly in 19 major producing states and monthly in 11 additional states. A Turkey Hatchery survey is conducted monthly in 16 states. Chicken and Egg surveys are conducted monthly to measure the number of eggs produced, rate of lay, number of layers and pullets and forced moltings. Major surveys for Aquaculture production (Catfish and trout) are conducted in January. Catfish processing survey is conducted monthly.

### *2.3.3 Environmental, Economics, and Demographics Statistics Program*

Environmental data are obtained from chemical and pesticide use surveys conducted for selected commodities on a rotational basis or every  $n^{\text{th}}$  year. Demographic data is obtained from the quarterly Farm Labor survey and the Census of Agriculture.

Annually, NASS conducts the Agricultural Resource Management Study (ARMS) to assess the economic well being of the farm economy. ARMS is conducted in 3 phases. Phase I conducted in May, screens farms to qualify for other phases. Phase II conducted September through December collects data on agricultural production practices, resource use, and variable costs of production, for specific commodities. Phase III conducted February – April, collects whole farm finance, operator characteristics, farm household information, operating expenditures, capitol improvements, assets, and debt for agricultural production.

A Prices Received survey is conducted monthly to estimate prices received by farmers for commodities. A Prices Paid survey is conducted annually to estimate farmer costs for inputs. Dairy product prices surveys are conducted weekly.

#### *2.3.4 Census of Agriculture*

The Census of Agriculture is conducted every 5 years. Special follow-on surveys include the Census of Horticulture, Census of Aquaculture, Farm and Ranch Irrigation survey, and an Agriculture Land Ownership survey.

### **3. PROCESSING CYCLE**

The Agriculture Statistics Programs being maintained by the three countries go through a series of processing steps before being released.

#### **3.1 Canada**

The Canadian Agriculture Statistics Program is based on a quinquennial Census of Agriculture and on seven major surveys that are conducted every year. Their major processing steps include the following.

##### *3.1.1 Frame Definitions and Exclusion*

The Census of Agriculture targets every household in Canada in order to identify agricultural operations and their operators. Up until 2006, the methodology was coordinated with the Census of Population. Enumerators delivered Census of Population questionnaires to all households and a Census of Agriculture questionnaire was left if it was determined that someone in the household operated an agricultural operation. In 2006 a subset of farms were mailed a Census of Agriculture questionnaire and the rest were delivered by traditional means. In 2011 all questionnaires will be mailed. In addition, the Census of Population questionnaire contained a question asking if anyone in the household was a farm operator, in order to find farm operators missed by enumerators. All operations had to return their completed questionnaire by mail or fill one out over the Internet. The resulting information was used to update the Canadian Farm Register.

During non-census years, tax records are used to update the Canadian Farm Register. The tax database is matched against the register and unmatched records which have reported farm revenue on their tax form are sampled, based on their Gross Farm Revenue, to complete a frame update questionnaire.

As part of the Canadian Agriculture Statistics Program, agriculture surveys use the up-to-date Farm Register as a list frame. Sampling frames generally exclude small farms, i.e. farms with less than 10,000 CAN\$ in agriculture revenue, as reported on the last Census. Adjustments are done at the estimation phase to take into account these excluded units and thus, to represent all operations with positive sales. Typically, these adjustments account for less than 1.0% of the major commodity estimates.

##### *3.1.2 Collection Steps*

The Census and surveys collect information for the agricultural operations. While the Census uses paper and internet collection methods, the surveys mainly use computer assisted telephone interviews (CATI) which take advantage of automated edits. All data collected by Statistics Canada correspond to specific reference periods or dates. Periods such as calendar year, month, crop season, etc. are used for flow data or farm practices. Dates are used for stock data or farm status at a specific point in time.

Typically, non-responses rates between 5% and 20% are observed for the Canadian agriculture surveys. These rates include refusal rates between 2% and 5%.

### *3.1.3 Data Editing*

Data editing is performed at the time of collection and later in batch as part of processing activities. Methods include a mixture of non-response edits, historical edits, consistency edits, statistical edits and outlier detection methods. All records go through the same set of automated edits. Typically, large units also go through manual edits.

### *3.1.4 Data Imputation*

When edit failures are observed, the erroneous fields are flagged for imputation. From a survey perspective, if the most important fields are flagged, then the record is deemed unusable for further processes thus reducing the set of respondents. The “usable” records may include edits failures. These are typically corrected with donor imputation methods when deterministic methods cannot be used. With this approach, imputation rates for key commodities are close to 0% for most Canadian agriculture surveys.

### *3.1.5 Estimation*

Survey estimates are produced by a regular Horvitz-Thompson estimator. Non-responding records along with other unusable records as a result of the editing phase are compensated through adjustments of estimation weights. A few outliers also benefit from weight adjustment techniques. Frame exclusions are adjusted at the macro level using census ratios.

Accuracy indicators in terms of variance and coefficient of variation are produced from the usual statistical theory.

Typically, coefficients of variation between 1% and 5% are observed for key commodities estimated by the Canadian agriculture surveys.

### *3.1.6 Disclosure Control*

Disclosure rules evaluate the direct and residual disclosure risks in tabulated cells prior to dissemination. Cells deemed to be at risk are suppressed.

### *3.1.7 Frame Updates*

Survey feedback is used to keep the Farm Register up-to-date, providing reliable information for subsequent surveys. Some Farm Register updates are survey specific, while others apply across all surveys.

## **3.2 Mexico**

The Mexican agricultural statistics program uses administrative registers to obtain most crop and livestock information; seasonal surveys for the production, prices and production costs of the majority of the important crops; and remote sensing to estimate the sown area of main crops. Also, a census of agriculture is conducted every ten years.

### *3.2.1 Administrative Data*

The state offices combine advanced reports on sown and harvested areas during the first ten days of each month and continue to capture this information in order to include the cumulated data in the next month's report. The central office provides the rules and the information system for processing the data. The main process for obtaining the administrative data is:

- 1) Information is captured in the systems of the SIACAP and SIPCAP programs by *CADERs* (Rural Development Support Centre) and/or *Distritos de Desarrollo Rural* (Rural Development Districts) in each office in the state.
- 2) The information is validated by the state offices.
- 3) The state offices send the data to the central office.
- 4) Computer support is given by the central office, which processes the information and produces the final database for consultation.

### *3.2. Sample Surveys*

The main activities that are carried out for the surveys are identifying the concepts and methods for gathering the correct information respond to the problem, designing the tools for collecting this information (questionnaires) and designing the results tables. The statistical design of the survey includes the following activities: identification of the observation unit, developing the sampling frame, calculation of sample sizes, establishment of criteria for selecting the sample, as well as criteria for calculating the estimates and logistics of the field work including the organization of field activities and the schedule for such activities. The area to be studied uses a list frame such as PROCAMPO (Program for Direct Assistance in Agriculture).

The central offices offer technical sessions for training technical personnel from the state on how to use questionnaires, and how they should be filled out, taking into account the two methods (surveys and the census) mentioned previously. In addition, the personnel coordinate the work and the activities to ensure that the methods are correctly applied. Other information is gathered from other sources to validate the information obtained through the questionnaires after it has been captured and stored in a database using software designed by the central office for that purpose. These surveys will obtain, by means of direct interviews with producers, information on production, target production, production costs and prices paid to producers as well as other products.

The state office carries out manual validation of the data contained in the questionnaires and electronic validation of the data contained in the databases through the use of edits. They also agree upon the value for a variable when different methods obtain different values. The office will carry out surveys to estimate harvests using interviews with producers, target productions, production costs, prices paid to producers and other surveys.

### **3.3 United States**

The U.S. Agricultural Statistics program is divided into two parts. The Agricultural Estimates program is a series of surveys and publications covering over 160 commodities and serving the business of agriculture. Surveys are conducted weekly, monthly, quarterly, semi-annually, annually, and bi-annually as determined through collaboration with each industry. The Census of Agriculture is conducted every five years and provides a detailed demographic and economic snapshot of American agriculture. Several follow-on surveys linked to the census are conducted in intercensal years.

#### *3.3.1 Frame Definitions*

All censuses and surveys start by defining the target population. In NASS, the main sampling frame is a **list frame** of farms and ranches with accompanying control data. The list frame allows NASS to efficiently sample farms and ranches for most commodities and many farm characteristics. For the ag estimates program, the list frame includes only those farms and ranches that have been confirmed as actively operating and qualifying under the farm definition. The Census of Agriculture targets every known farm operation in the United States maintained on NASS's universe list of farms and ranches. In contrast to the ag estimates list frame, the Census Mail List (CML) also includes potential farms whose operating status may not be confirmed. While new list sources are introduced continually, an intensive list building effort is conducted prior to each census. A significant problem with the list frame and the CML is that they are incomplete. The list frame is presumed complete for many small specialty commodities and for all agribusinesses.

NASS has also built an **area frame** based on all land in the United States and, consequently, includes all farms. The area frame is divided into segments of land within strata defined by degree of cultivation. The area frame is complete; however, a general purpose sample of segments is not efficient for many commodities. NASS marries these two frames using multiple frame sample design to utilize the efficiencies of the list frame and use the area frame to measure incompleteness. Nearly all major sample surveys conducted by NASS utilize a multi-frame sample to measure the incompleteness of the list frame. It is absolutely imperative that the List and Area Frames remain statistically independent.

#### *3.3.2 Data Collection Steps*

NASS uses all modes of data collection, mail, telephone, face to face, and Internet. Most ag estimates surveys utilize a combination of modes with most data collected by telephone. The Census is administered via mail with telephone nonresponse follow up and limited personal interviewing of larger farms. Computer assisted telephone interviews and web-

based survey instruments allow for basic, real time editing. Area frame surveys are conducted via personal interview only. The ARMS, a large economic study, is conducted via personal interview, with an optional shorter mail version instrument for selected respondents. All surveys conducted by NASS have a defined reference date. This reference date may be a specific day or an extended period of time. For example, expense and income data cover a full year while prices received for crops are usually for a whole month. Acreages and final yield estimates refer to the growing season. Items that are constantly changing, such as livestock inventories and yield forecasts, are measured as of a specific day. All NASS publications clearly state the reference period.

### *3.3.3 Data Editing*

Data editing may be performed at the time of collection and in batch as part of the processing activities. Methods include a combination of nonresponse edits, historical data checks, consistency edits, statistical edits, and outlier detection methods. Paper versions of the survey instruments typically go through manual edits in addition to the machine edits.

### *3.3.4 Data Imputation*

NASS has several imputation strategies at its disposal for dealing with item and unit nonresponse. For a given survey, the most appropriate strategy(s) is chosen. Some imputation strategies make the data good by supplying a value while others adjust the sample weight. Some strategies use a reporting farm of similar size and type and others form imputation groups of similar size and type. Automated imputation algorithms are based on the fundamental characteristic of homogeneity, regardless of whether individual farms or groups are used. Very large or unique farms usually do not fit homogeneous groups or have a similar farm to draw from. These farms must be manually imputed by statisticians.

### *3.3.5 Estimation*

Most survey summaries produce multiple point estimates, often a combination of direct measures and ratio estimates. Standard errors and coefficients of variation are calculated using usual statistical theory. Summaries are run at the state level and, for national surveys, state results are aggregated to the national level.

Field Offices are responsible for doing a thorough data review, interpreting survey results, and submitting recommendations to Headquarters. NASS employs a top down approach for establishing official estimates, setting the national estimate and adjusting selected state recommendations to ensure states add to the national total.

Surveys are designed to achieve coefficients of variation of 1 to 3 percent for U.S. estimates and 5 to 10 percent for state level estimates for key commodities.

### *3.3.6 Disclosure Control*

Federal law requires NASS to withhold any estimate that would disclose individual farm data or allow the public to closely approximate a farm's attributes. NASS uses a two part disclosure rule to determine primary suppressions. A threshold rule requires three or more farms produce the commodity before it can be published. A dominance rule protects very large farms from having their data revealed. The ag estimates program uses the (n,k) rule to determine dominance which suppresses a value if the largest n farms produce k percent or more of the total. The Census program uses the p-percent rule which suppresses the value if the residual, after removing the two largest farms, does not provide p percent protection for the largest farm. NASS does not disclose the parameters used in either dominance rule. An estimate failing one or both of the disclosure rules is called a primary suppression. Often a primary suppression is part of a published total and a second (complementary) suppression is required to protect the primary.

### *3.3.7 Frame Updates*

Updates to the NASS List frame and Area frame are done on a continuous basis as part of standard survey procedure. Typically, updates are completed immediately following the completion of a survey.

## **4. ADJUSTMENT METHODS**

During the estimation phase, the Agriculture Statistics Programs being maintained by the three countries need to account for undercoverage and other issues prior to release.

### **4.1 Canada**

#### *4.1.1 Coverage*

The coverage of Canadian surveys highly depends on their frames and therefore on the Census of Agriculture which feeds the Farm Register. In spite of efforts by census representatives to locate and enumerate all farm operations in Canada, the Census of Agriculture misses some farms, primarily because of the difficulty in correctly identifying an agricultural operation when none of its farm operators live on or near it. To reduce undercoverage, census representatives are instructed to ask a member of every household whether someone in the household is a farm operator. In addition, since 1991, an agriculture operator screening question has been on the Census of Population questionnaire to identify farm operators missed when the questionnaires were delivered. If a Census of Population questionnaire was returned with this question marked "yes," the Missing Farms Follow-up Survey could call those households by telephone to complete a Census of Agriculture questionnaire.

The Coverage Evaluation Survey was designed to survey households that were not called by the Missing Farms Follow-up as well as Farm Register operations for which no questionnaires were received during the activities of the Census of Agriculture. A score function was applied to all units to identify their priority for the coverage survey. The targeted units were then stratified based on their score and a sample was drawn. The survey

used a very short questionnaire to collect only key variables. As a result, the 2006 Census showed a 3.4% undercoverage rate of farms and undercoverage rates below 2.0% for major commodities.

As part of the Agriculture Statistics Program, survey results are benchmarked against the censuses and so, the Census of Agriculture coverage results are recycled for the surveys.

#### *4.1.2 Other Adjustments for Concepts*

As mentioned in 3.1.1, frames target all operations with positive agriculture revenue although operations with less than 10,000 CAN\$ in revenue are excluded from sample selection. These operations are rather accounted for by adjustment factors, with no error measures.

## **4.2 Mexico**

### *4.2.1 Coverage*

Coverage from the thematic point of view corresponds to the agriculture, livestock, aquaculture and fishing sectors.

From the point of view of type of production unit (all goods and services and land managed under the same administration), all units are included, no matter the size of the property, type of ownership (communal, private or public), nor amount of revenues.

As of 2009, SAGARPA (Secretary of Agriculture, Ranching, Rural Development, Fisheries and Food Supply) is developing an information system based on objective methods such as sampling and geomatics, using materials and results from the agriculture and livestock census as inputs. This will be done for the main crops and production zones; for the remaining crops, it will be supplemented with the administrative-type information so as to ensure complete coverage.

### *4.2.2 Other Adjustments*

The two systems (administrative records and censuses) include the same number of producers being studied, that is to say there are no differences between the groups.

Therefore, there is no adjustment method required to account for under- or over-estimation.

## **4.3 United States**

### *4.3.1 Coverage*

Coverage adjustments measures are used to address list incompleteness and to ensure the adequate representation of the responding sample. List incompleteness is, by far, the greater challenge and NASS builds incompleteness measures into its surveys and censuses.

Each year in June, NASS performs a complete enumeration of the area frame sampled segments and determines who operates farms in the segment. The names are checked against the list to determine which ones are in our sampling populations (overlap) and which ones are not (non-overlap). Non-overlap farms are surveyed to directly measure what is not covered by the list. Representativeness is a consideration for surveys and censuses that cover multiple agricultural industries (cattle, field crops, fruit). After nonresponse and under-coverage adjustments are made, some industries may be over represented while others come up short, resulting in biased point estimates. Coverage adjustments are achieved by calibrating sample weights to meet externally provided targets. Procedurally, calibration finds a solution to a system of linear equations (constraints) defined by the targets

#### *Agricultural Estimates*

For all national surveys, list incompleteness is directly measured using non-overlap tracts. For crop and livestock surveys, the summaries treat the sample as representative and no adjustments are made. For the annual economic survey, weights are calibrated based known production targets.

#### *Census of Agriculture*

CML incompleteness targets are defined by adding a directly measured count of missing farms to the count of reporting census farms adjusted for nonresponse for over 60 variables. Production targets, supplied from non-census sources, are used to assess representativeness. Both sets of targets are merged and calibration adjusts the weights. Thus, incompleteness and representativeness are addressed concurrently within each state.

#### *4.3.2 Imputation*

##### *Agricultural Estimates*

For livestock surveys, the edit identifies cells requiring item imputation and refers them to an analyst for manual resolution. Unit nonresponse is performed by reweighting within groups. Livestock surveys are traditional stratified designs so weighting groups are imbedded in the design. Acreage and Production surveys use a custom algorithm for item and unit nonresponse for key commodities that imputes for missing data. The algorithm uses weighting groups based on size and type subdivided by geographic region to create more homogeneous groups for cropping practices and crop yields. The annual economic survey uses weighting groups by size and type. Groups falling short of a minimum sample size requirement are combined with other groups until a sufficient number of reports is achieved.

##### *Census of Agriculture*

Item imputation is completed by making the data good. When missing data cannot be derived deterministically from other reported data and the item has not been previously reported on another NASS survey, a “nearest neighbor” algorithm is invoked. A similar

farm is identified as the Euclidian distance between the recipient and prospective donors (clean reports) using selected variables on the CML. Match variables are specific to each section of the report form and include the latitude and longitude of the principal county of operation. The donor with the smallest distance is considered the nearest neighbor and becomes the source for the imputation action. Since imputation is conducted independently for each occurrence, reports requiring multiple imputations draw from multiple donors. Initial donor pools are established before the first batch edits were run. These donor pools are seeded with previous census data “mapped” to look like the current census data and passed through the current edit plus data collected during the current census content test. If the imputation algorithm fails to provide an acceptable solution, the record is referred to an analyst for resolution.

Whole farm nonresponse adjustments are made using modeled estimates of in-scope farms within weighting groups at the state level. By design, the CML contains many more names than there are qualifying farms and, thus, many nonrespondents do not qualify as farms. Data mining techniques are used to define weighting groups. A stepwise approach is used to define weighting groups. All farms on the CML are categorized based on a predetermined list of defining variables. Tests of significance are performed on the response rates in each group. If significant, the group is permanently set and the next variable is tested for the names in each group. If not significant, the names are recombined and the next variable is tested. When all variables have been tested and the final weighting groups are defined, the in-scope rate of the respondents is calculated and applied to the count of nonrespondents to estimate the number of CML farms not accounted for. The weights of the respondents in the group are scaled to represent the missing farms.