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 biennial, annual 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, (c) the biennial Farm Financial Survey conducted in the summer to collect information about revenues, expenses and properties, as well as physical characteristics, (d) the Fruit and Vegetables Survey conducted in November to collect area and production of fifty different fruits and vegetables, and (e) the annual Greenhouse, Sod and Nursery Survey conducted in January to obtain a view of the general economic situation within these industries. Smaller surveys include (f) the Potato Area and Yield Survey, (g) the Hay and Straw Prices Survey, (h) the Corn and Soybean Survey, (i) the Mushroom Survey, (j) the Honey Survey, (k) the Maple Products Survey, and (l) the Production of Poultry and Eggs Survey.

In addition, a Farm Update Survey (FUS) is conducted regularly, principally to verify potential births to be added to the Business Register and to the survey frames. This survey makes use of tax return forms and administrative lists to identify units with a high probability of being new farms.

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 and the Agricultural Water Survey are examples.

The Agriculture Statistics Program is completed by the Tax Data Program, the Farm Income Statistics Program, Remote Sensing and Geospatial Analysis 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 conduced 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 corn, soybeans, wheat, cotton, and potatoes, in major producing states. Other surveys varying in frequency include the Floriculture survey and the Citrus Fruit survey. Non-Citrus Fruit, Nut and Vegetable Forecasts and Estimates were suspended in 2013.

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 including Catfish Feed Deliveries and Catfish Processing), July Cattle Report, and Mink Report were suspended in 2013

2.3.3 Environmental, Economics, and Demographics Statistics Program

Environmental data from chemical and pesticide use surveys conducted for selected commodities on a rotational basis or every n^{th} year were suspended in 2013. Demographic data are obtained from the semi-annual 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. In 2006 a subset of farms were mailed a Census of Agriculture questionnaire and the rest were delivered by enumerators if it was determined that someone in the household operated an agricultural operation. In 2011 all questionnaires were mailed. In addition, the Census of Population questionnaire contained a question asking if anyone in the household was a farm operator, in order to identify missing farms. All operations had to return their completed questionnaire by mail or fill one out over the Internet. The resulting 2011 information was used to update the Canadian Business Register.

During non-census years, the Canadian Business Register uses tax records to identify new potential farms. Records which have reported farm revenue on their tax form are sampled, and asked to complete a frame update questionnaire to determine if they represent a farm and what commdities they raise.

As part of the Canadian Agriculture Statistics Program, agriculture surveys use the up-to-date Business 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 then 1.0% of the major commodity estimates.

3.1.2 Sample Design and Selection:

The Census of Agriculture uses all of the active units found on the Canadian Business Register for the initial mailout of the Census questionnaire.

The sample design for individual agriculture surveys starts by identifying all of the farming operations in the population of interest according to the information on Canadian Business

Register. This population is stratified by geography, size and in some cases, farm type. The smallest farms are placed into a take-none stratum and are not eligible to be selected for the sample. Using information from the frame related to the variables of interest, the sample is allocated using either univariate or multivariate allocation methods so that the expected survey estimates for these key variables meets a prescribed precision. Permanent random numbers are assigned to each unit in order to control the amount of overlap in the sample from one survey occasion to the next.

3.1.3 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%.

3.1.4 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.

3.1.5 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.

3.1.6 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.

3.1.7 Disclosure Control

Disclosure rules evaluate the direct and residual disclosure risks in tabulated cells prior to dissemination. Direct disclosure may occur because either the cell has very few non-zero contributors to the estimate or a very small number of contributors may account for a very large percentage of the overall estimate. Automated methods are used to identify those cells which are in direct risk of being disclosed. Manual methods are used to control the residual disclosure risk. In both cases, cells deemed to be at risk are suppressed. Care is taken to consider tables which have been previously released in order to avoid residual disclosure between more than one table.

3.1.8 Frame Updates

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

3.2 Mexico

3.2.1 Frames definitions and exclusions

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

3.2.1.1 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 *CADER*s (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.

3.2.1.2 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.

3.2.2 Sample design and selection

The surveys use all of the active units found in the Agricultural census plus the administrative registers as a frame sampling list. Also, local information from governmental programs like the producers list of ASERCA is used. ASERCA is a federal government entity that gives subsidies to agricultural producers, and so it has a list of the supported producers and other key variables that are used in the building of list frames.

The sample design for individual agriculture surveys starts by identifying all the operations in the population of interest according to the information of the available sources and creating a single list. The population is stratified by geography, production and crop performance. All strata are eligible.

Using information from the frame related to the variables of interest, the sample is allocated using multivariate allocation methods, so that the expected survey estimates for the key variables meet a prescribed precision.

Random numbers are assigned to each unit in order to control the overlap in the sample.

Additional samples may be drawn from populations with special characteristics. In these cases the weights are corrected in a post stratification process.

3.2.3 Collection steps

The surveys collect information for the agricultural operations by paper and internet collection methods. The information is used to estimate flow data or farm practices for a specific period such as calendar year, month, crop, season, etc.

Typically non responses rates are between 10% and 20% these rates include refusal rates between 5% and 7%

3.2.4 Data Editing

Data editing is performed at the time of collection and later in batch as part of processing activities. These edits verify that critical fields have been filled in, ensure that the data is internally coherent and apply outlier detection methods. This is done by manual editing and in some cases automatic edits.

3.2.5 Data imputation

When edit failure is observed, the erroneous field is flagged for imputation. A correction and verification process is held with the person that captured the information if possible. If the problem persists the record is deemed unusable for further processes.

3.2.6 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.

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

3.2.7 Disclosure Control

Disclosure rules evaluate the direct and residual disclosure risks in tabulated cells prior to dissemination.

Direct disclosure may occur because either the cell has very few non-zero contributors to the estimate or a very small number of contributors may account for a very large percentage of the overall estimate.

Automated methods are used to identify those cells which are in direct risk of being disclosed.

Manual methods are used to control the residual disclosure risk. In both cases, cells deemed to be at risk are suppressed. Care is taken to consider tables which have been previously released in order to avoid residual disclosure between more than one table.

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 designs 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 an area 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 Sample design and selection

NASS employs several different sample designs. For small target populations, like mink producers and vegetable processors, a census is performed. For single commodity surveys, such as livestock, stratified designs are used with simple random sampling within strata. Surveys covering multiple commodities, such as crop acreage and grain stocks, multivariate probability proportional to size designs are used. Probabilities of selection are computed for each commodity of interest present on a farm's control data profile. PPS samples are

selected for each commodity and each selected record is initially given a weight inversely proportional to its maximum probability of selection. Sampling weights are calibrated to the frame totals for the commodities of interest. For some surveys, permanent random numbers are used to control respondent burden.

3.3.3 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, computer assisted personal 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.

3.3.4 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.

3.3.5 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.

3.3.6 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, adopting 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.

3.3.7 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.

3.3.8 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 or census.

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 Business 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, 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.

Unlike previous Censuses of Agriculture, the 2011 Census did not have an explicit Coverage Evaluation Survey designed to survey households that were not called by the Missing Farms Follow-up as well as Census frame operations for which no questionnaires were received during the activities of the Census of Agriculture. Instead an estimated undercoverage rate was calculated based upon the observed results of the Missing Farm Follow-up process. As a result, the 2011 Census showed a 1.8% undercoverage rate of farms and undercoverage rates below 1.0% for total farm area and total farm sales.

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 adjustment 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.

5. QUALITY ASSURANCES PROCESSES AND DATA QUALITY INDICATORS

Throughout the various phases of the census or survey process quality assurance methods have been implemented. In addition data indicators are used to quantifiably describe the quality of the data or the process.

5.1 Canada

5.1.1 Data Quality Assurance Practices

All surveys and censuses have some degree of quality assurance. In the case of the Canadian Census of Agriculture there are well-defined, specially developed quality

assurance plans in place for the most important steps. In other surveys more general practices are used.

5.1.1.1 Questionnaire Design

Statistics Canada's Questionnaire Design Resource Centre is used to review and test any new questionnaires or questionnaires which have undergone significant change. This process may include a simple review of the questions or a more complex process of testing the questions, their ordering and their flow on a sample of respondents in a focus group setting. A report on the Centre's findings is provided to the survey mangers including suggested changes if necessary.

5.1.1.2 Sample Frames

The surveys and Census of Agriculture have a mutually beneficial relationship for maintaining the quality of the survey frames. The Census provides a complete listing of units which is used as the starting point for the survey frames. The surveys keep this frame up to date and relevant through survey feedback. A Farm Update Survey also uses information from external sources such as producer associations to identify any potential new farms. A survey is then sent to these farms to gather information on their activities. Based on this information the farm may or may not be added to the frame.

5.1.1.3 Sample Design, Selection and Response Burden

Because the agriculture surveys are activity specific, the survey designs can be made very efficient by making use of frame information on those activities. This leads to accurate estimates with a reduced sample size. Statistical methods are used to control the amount of sample overlap between and within surveys to reduce response burden. For most surveys, Statistics Canada generalized software tools are used for sample selection

5.1.1.4 Data Collection

Interviews for surveys are conducted from Statistics Canada's regional offices. Interviewers are specifically trained in agriculture surveys. In addition, specialized interviewers are used to conduct more difficult cases such as the conversion of people who refuse to participate. An integrated collection system that imports the sample, assigns individual cases to interviewers and sends the collected information back to Head Office is used. Information on survey progress is provided to regional office managers on a daily basis to help them to manage the workload. Most surveys are collected using a Computer Assisted Telephone Interviewing application which has important edits embedded within. Thus errors and inconsistencies can be identified during the interview and resolved immediately. For more complex surveys, a copy of the questionnaire is mailed to the respondent ahead of time so that he/she may review it and prepare the answers ahead of time.

5.1.1.5 Data processing

Quality control procedures are in place for data that needs to be manually captured or coded. An agriculture-specific system is used for the processing of nearly all of the surveys. Analysts are able to review individual records and compare the reported data with that from previous survey occasions. Imputation is used for partially completed records. This is done either by the analysts' knowledge or by automated methods. A number of automated imputation methods including historical, donor, mean and other methods are available.

5.1.1.6 Estimation and Dissemination

The weighting and estimation is done using well-accepted approaches. Auxiliary information from the Census is used to adjust the survey estimates to account for the non-surveyed population. Comparisons are made between survey results and other available data sources such as producers' association estimates. Documentation on the methods and approaches used is provided with each topic so that the user understands how the estimates were generated.

5.1.2 Data Quality Indicators

Quantitative data quality indicators are generally produced at two stages in the production cycle – during the data collection and at the processing and estimation stage

5.1.2.1 Quality Indicators Related to Data Collection

Response rates are used throughout the survey process. As the collection progresses, they are monitored and compared to the response rate targets. In cases where the observed rates are significantly lower than the targets, the collection period may be extended. Two types of response rates are calculated – either unweighted or weighted based upon a value on the Business Register related to the survey variables of interest. A distinction is also made between collection and final response rates. The former represents the percentage of cases considered to be respondents following the data collection, while the latter represents the percentage of respondents after the data has been cleaned and processed.

Reasons for non-response are also monitored over time. An increase in the amount of a certain type of non-response such as refusals may result in revisions to the instructions given to interviewers. They are also used to perform proper adjustments to the sampling weights to account for non-respondents

In the Census the coverage error is determined through the results of the Missing Farm Follow-up process as described in section 4.1.1. Estimates of the amount of undercoverage are determined for not only the number of farms, but also for a number of major commodities.

Paradata, or data related to the collection of data itself, is increasingly being used to provide indicators related to the data collection process such as the length of time that it takes to complete an interview or the amount of burden being put on the respondent.

5.1.2.2 Quality Indicators Related to the Estimates

During the design of the sample a set of precision targets are established by important variable of interest. These targets are expressed in the form of coefficients of variation (CVs). Sample is allocated in such a manner to minimize the number of units to be contacted while at the same time respecting these targets. During estimation, the true surveys CVs are calculated. The observed CVs are compared to the theoretical ones and important differences are investigated to explain them. CVs of individual estimates are often published using a letter rating rather than the actual CV value. A letter from A to F is assigned to each estimate which indicates the quality of the estimate.

Main estimates are compared from one survey cycle to the next. Those which have large differences are investigated to determine the reason. Estimates are also compared to those from other sources of data, such as survey, Census or administrative sources. Large differences are investigated.

5.2 México

5.2.1 Data Quality Assurance Practices

All surveys and censuses have some degree of quality assurance. Some general practices are used like internal coherence and statistical filters, but they depend on the survey.

5.2.2 Data Quality Indicators

Quality control procedures are in place for data that needs to be manually captured or coded. An agriculture-specific system is used for the processing the surveys.

5.2.2.1 Quality Indicators Related to Data Collection

No quality indicators are specially used or recorded in this process just the number of records that were consigned and modified.

Response rates are used throughout the survey process.

Two types of response rates are calculated – either unweighted or weighted based upon a value on the frame list related to the survey variables of interest.

A distinction is also made between collection and final response rates. The former represents the percentage of cases considered to be respondents following the data collection, while the latter represents the percentage of respondents after the data has been cleaned and processed.

Reasons for non-response are also monitored over time. An increase in the amount of a certain type of non-response such as refusals may result in revisions to the instructions given to interviewers. They are also used to perform proper adjustments to the sampling weights to account for non-respondents

5.2.2.2 Quality Indicators Related to the Estimates

During the design of the sample a set of precision targets are established by important variable of interest.

These targets are expressed in the form of coefficients of variation (CVs). Sample is allocated in such a manner to minimize the number of units to be contacted while at the same time respecting these targets.

Main estimates are compared from one survey cycle to the next. Those which have large differences are investigated to determine the reason. Estimates are also compared to those from other sources of data, such as survey, Census or administrative sources. Large differences are investigated.

5.3 United States

5.3.1 Data Quality Assurance Practices

Quality assurances measures are implemented for all surveys and census conducted by NASS. A set of policy and standards are established to ensure consistent, correct application of procedures for frequently occurring activities. A standard is established for any statistical activity that occurs frequently under similar conditions – either across time, across surveys, or across organizational units.

5.3.3.1 Questionnaire Design

Pre-testing of survey instruments is performed to ensure that each question within every survey instrument effectively and accurately captures the data appropriate to the estimate(s) it supports.

All survey instruments for the census and other federal surveys are coordinated by the Census and Survey Division at NASS Headquarters and disseminated to State Field Offices. Questionnaire design and testing is coordinated by survey methodologists in NASS's Research & Development Division. CATI and CAPI instruments, and web based questionnaires provided to the Field Offices must be consistent with the paper questionnaire for each survey. Differences which exist between modes (self administered, face-to-face, and telephone) of data collection are considered in the design and content of the respective Questionnaires. All paper, web-based, and computer assisted survey instruments for federal surveys are subject to review and approval by the U.S. Office of Management and Budget.

A number of major surveys go through a formal specifications process to address requests for changes to wording or adding or dropping questions.

5.3.3.2 Sample Frames

NASS is charged with the task of maintaining an up-to-date list of farm operations within the United States, along with indications of size and type of operation, referred to as control

data, which are important for development of efficient sample designs. **All types and sizes of agricultural operations** must be represented on the list. In order to assure the most effective use of resources in pursuing this task, NASS sets priorities for establishing coverage for different subgroups, of the agricultural sector.

The List Frame Council, Statistics Division, and the office of Field Operations consult to establish list frame coverage goals at the national and state levels, as appropriate, for all farm operations, important demographic and farm sector subgroup and production of important commodities.

List frame records representing current agricultural establishments with agricultural activity reported on a recent census or survey instrument are maintained as an active record on the list frame. This includes records that may not meet $1,000 in sales criteria but report agricultural activity.

5.3.3.3 Sample Design, Selection and Response Burden

NASS uses several sampling and estimation procedures for reducing the incidence of multiple contacts for sampled operations. Sequential Interval Poisson (SIP) sampling is used to select samples for the Agricultural Resources Management Survey (ARMS) Phase III sample to minimize overlap between the current year ARMS sample, the previous year's ARMS III sample and the current year Quarterly Agricultural Survey sample. Sequential Interval Poisson (SIP) sampling is also used to avoid overlap between the Small Grains Agricultural Yield Survey sample and the Row Crops Agricultural Yield Survey sample in the Quarterly Agricultural Survey Program.

Replicated Sampling has traditionally been used to control overlap of samples drawn from the same population. Permanent Random Number (PRN) methodology is used to assign each eligible sampling population member a PRN for some surveys. Samples can be selected in PRN ranges to control the overlap of samples. Non-Overlap with the List (NOL) modeling uses reported data from list operations of comparable size to model the NOL portion of the estimator thus avoiding contact with NOL operations during subsequent survey periods for multiple frame surveys.

All surveys with overlapping data collection periods are coordinated such that data collection is completed during one contact, unless the operator requests separate contacts. Non-monthly surveys having reference dates one month or less apart are coordinated for single contact data collection to be made as late as possible within the earlier survey's data collection period.

5.3.3.4 Data Collection

Data collection activities for most federal surveys are coordinated through NASS's Census and Survey Division. Data collection is carried out by State Field Offices and Regional Data Collection Centers.

The National Association of State Departments of Agriculture (NASDA), through a cooperative agreement with NASS, provides a qualified enumerator corps to collect agricultural data. The goal of both organizations is to conduct a practical and effective training program for NASDA employees. An effective training program will contribute to consistent data collection methods, a minimum of data collection errors, and high quality agricultural statistics.

The NASS/NASDA policy is enumerators will perform data collection activities only after being properly trained. The responsibility for implementing general and survey-specific training rests with the Field Office (FO) NASDA coordinator (Senior Survey Statistician) and supervisory enumerators.

5.3.3.5 Data Processing

A SAS based Generalized Summary System (GenSumm) aggregates survey data, calculates precision levels, provides supplemental analytic information, and displays the results in an easy to read format. It is based on the SPS Summary System which has been used by NASS since the early 1990's. Both tools use the same underlying programs to calculate the indications and associated data, providing consistent, statistically sound analysis.

5.3.3.6 Estimation and Dissemination

Commodity estimates are developed through a predetermined statistical process, beginning with data obtained from surveys of producers and agribusiness firms. All survey data are validated through a detailed editing and analysis process. Where possible, these survey data are compared with administrative data from various sources and knowledge gained through travel and frequent personal contact with individuals involved in the production, marketing, and administration of agricultural commodities to insure the data published are reasonable, reliable, and reflect current trends in agriculture.

With few exceptions, data are collected, edited, analyzed and summarized at the State level. State estimates are prepared for submission to the Agricultural Statistics Board (ASB). The Board, a panel of commodity experts, reviews the combined and individual State recommendations, and adopts and publishes State, regional, and national estimates.

5.3.4 Data Quality Indicators

Quantative data analyses are produced during data collection, processing, and estimation. The U.S. Office of Management and Budget (OMB) requires quantitative analysis of all federal survey activity that is subject to approval by OMB.

5.3.4.1 Quality Indicators Related to Data Collection

For every data collection effort, one or more quality assurance procedure(s) are utilized. Statisticians and supervisory enumerators are responsible for quality assurance through: (1) re-contacting a subset of respondents to verify data, (2) reviewing some or all completed paper questionnaires for proper completion, flow and notes, and (3) monitoring interviews

to verify that enumerators are using effective interviewing procedures and to determine where additional training is needed. For monitoring purposes, field enumerators are periodically accompanied by a supervisory enumerator on one or more interviews. Telephone enumerators are observed via electronic equipment which permits the supervisor to monitor the CATI screens and conversation (where available) to are assure interviewers following established procedures.

The U.S. Office of Management and Budget sets required target response rates for surveys conducted by federal agencies. Response rates and coverage adequacy are monitored throughout the data collection process. Response rate calculations are both weighted and unweighted.

Detailed analysis of response rates between survey periods and over time are maintained and documented in reports for each major probability survey program. These detail analysis are performed at the state and national level and include reasons for non-response such as refusals or inaccessibles, non-response by modes of data collection, and non-response by respondent type (operator, spouse, bookkeeper, etc).

A coverage evaluation survey is conducted for the Census of agriculture to provide a measure of coverage error. Estimates of the amount of undercoverage are determined for number of farms, major commodities, and selected demographic categories.

5.3.4.2 Quality Indicators Related to the Estimates

NASS establishes standards for statistical precision of the indications from major national probability surveys. The standards for statistical precision of the indications are expressed as coefficients of variation (CVs).

Precision levels are established for specific survey indications. Six factors are considered in determining what the target values should be. These factors were (in no particular order): 1) expected use of the estimate by data users, 2) impact on Agency strategic performance measures, 3) compatibility between region/State targets and the U.S. targets, 4) relative magnitude of the estimate, 5) achievability given agency appropriated resources, workload, response burden, and quality of sampling information, and 6) historical experience.

Responsibilities for evaluating the performance of each survey are assigned to appropriate agency units. The individual components of the CV (i.e., the list and area frame components) are evaluated and recommendations made to enable each survey to regularly meet the target CV, including sample design modifications and sample size changes.

Last updated: June 2013