

QUALITY APPROACH TO AGRICULTURE CENSUS TAKING

Elizabeth Ann Vacca and Ruth E. Detlefsen, U.S. Bureau of the Census
Elizabeth Ann Vacca, U.S. Bureau of the Census, Washington, D.C. 20233

Key Words: Census, Quality Approach

1. What is the Census of Agriculture?

The Bureau of the Census is mandated by law to conduct a census of agriculture every five years to provide data on agricultural operations for each state in the U.S., the Virgin Islands, Guam, the Commonwealth of the Northern Mariana Islands, the Commonwealth of Puerto Rico and, as determined by the Secretary of State, other possessions and areas over which the U.S. exercises jurisdiction, control, or sovereignty. The census of agriculture is the leading source of statistics about the nation's agriculture and the only source of consistent, comparable data about agriculture at the county, state, and national levels. The census of agriculture strives to enumerate all places from which \$1,000 or more of agricultural products were sold (or normally would have been sold) during the census year. The census is conducted every five years in years ending in 2 and 7, the most recent collection being for 1992. Since 1969, the census of agriculture has been conducted using a self administered mail data collection procedure in lieu of personal enumeration. The primary goal of the census of agriculture is to provide timely, high quality data on all farms and ranches and on all farm production activities, for every county in the nation. The census of agriculture is a rich source of detailed information about farm and ranch operations.

2. What is a Quality Approach to Census Taking?

A quality approach to census taking is a commitment to excellence throughout the census process. There are many steps in the successful implementation of a census. Objectives must be defined, questionnaires designed and printed, a mailing list constructed, necessary samples designed and selected, methods and equipment for data collection and processing determined, questionnaires mailed, data collected, nonresponding addresses followed up, data reviewed and edited, missing values imputed, estimates computed, tabulations produced and results published. The approach that ensures quality builds quality into each step of the process. Typically, quality is measured in terms of

sampling errors after the census has been conducted. However, much can be done prior to and during the conduct of the census to control nonsampling error. A comprehensive look at the agriculture census illuminates a quality approach to census taking.

3. How is Quality Assured in Agriculture Census?

Quality is assured in the Agriculture Census by careful planning; conducting pre-census tests of questionnaires, methodologies and materials; creating a high quality mail list; control and measurement of data collection and processing operations; researching and development of statistically sound design and estimation methodologies; production of statistical measures for user information; and post-census evaluations for continuous improvement.

3.1 Careful Planning - A critical factor in a quality approach is the careful planning of the processes and interfaces linking the processes. This is where quality begins for the agriculture census. Prior to each census, a team of survey statisticians, computer specialists, and mathematical statisticians representing middle and upper management spend many hours delineating the major census processes and assigning target completion dates to each process. This provides a navigational guide and schedule to track and monitor progress throughout the census. The involvement of management demonstrates their commitment and ensures the goals and objectives are clearly defined and understood. The team stays functional throughout the census to facilitate communication, monitor progress and resolve problems that may arise.

3.2 Pre-Census Tests - The Census Bureau has an ongoing questionnaire design research and testing program. The tests are designed to evaluate alternative methodology and procedures and the question wording of new data items. The tests ensure that proposed changes are indeed needed improvements and that quality data from new

¹This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.

questions can be obtained. Purposes of proposed changes are to achieve higher response rates, reduce respondent burden, incorporate current data needs, increase the timeliness of the census, and introduce efficiency into the census process. Tests are conducted to assess changes to questionnaires, content, instructions, letters, mail follow-up patterns, and Computer Assisted Telephone Interview (CATI) instruments. The quality of the census statistics is affected by many factors. Among these are the effectiveness of the data collection instrument in eliciting response, the comprehensibility of the questions and instructions, and the completeness and correctness of reporting of individual data items by respondents.

3.3 Mail List Development(MLD) - Quality of the census published data is affected by the quality of the mail list. A high quality mail list provides complete coverage of the farm and ranch population and no coverage of the non-farm population.

The mail list is developed by assembling records of individuals, businesses, and organizations identified as having some association with agriculture. Records come from the previous census, the Internal Revenue Service, the Social Security Administration, and the U.S. Department of Agriculture. Additional lists are obtained for specialized operations (e.g. nurseries and greenhouses, specialty crop farms, poultry farms, fish farms, livestock farms, cattle feedlot operations, grazing permits) from State and Federal government agencies, trade associations, and similar organizations.

Although the mail list is compiled from lists of addresses associated with agriculture, not every record represents a farm operation. A large number of addresses contain an unknown farm status. The mail list development process is a balance between maintaining coverage of the farm universe and eliminating duplicate and non-agricultural related addresses from the list. Due to unknown farm status and to difficulties in identifying exact matches, it is not possible to limit the mail list to only those addresses certain to be farm operations or to eliminate many potentially duplicated addresses. Doing so would adversely affect list coverage. The census matching procedures have been designed to identify almost exact matches. To a large degree we rely on questionnaire recipients to identify duplication. This increases coverage at the cost of duplication. While ensuring a perfect mail list for the census of agriculture is an impossible task, special efforts are made to move as close as possible to a complete and accurate list.

Five record linkage operations are used to remove

duplicate addresses from the list. The addresses on the lists are formatted and standardized, geographically coded and ZIP code verified, linked using business and personal identification number and alphabetic name and address, and clerically reviewed if determined to be potential duplicates.

Many specific operations are used during the above operations to increase the accuracy of the address and agricultural data on each record. The accuracy of the address information is critical because it affects the ability of the Census Bureau to find the addressee and thus lower mail and follow-up costs. The accuracy of the agricultural data, which gives an indication of the size of each operation, contributes to proper classification of an address as eligible for sampling or not.

Once the initial linkage and unduplication process has been completed, we take steps to further reduce the mailing list by removing nonfarm addresses and additional probable non-farms from the list. Administrative record information, past census information and statistical modeling is used to determine probable farm status. Statistical modeling is performed using classification analysis to separate mail list records into groups according to the proportion of expected census farms. The groups of addresses least likely to represent farms are removed from the mail list.

The final 1987 and 1992 census mail lists contained 4.1 million and 3.6 million addresses of which 2.1 and 1.9 million were estimated to be farm and ranch operations, respectively.

3.4 Census Sample Design and Estimation - Quality data is most often judged by the soundness of the statistical methodology and the values of the sampling errors. Sampling is used in the census to reduce respondent burden and to estimate for nonresponse. The basic agriculture data about a farm operation such as acreage, land use, crops and livestock on hand, sales information, etc., are collected from all farm operations. A sample of approximately 25 percent of all mail list records are asked to provide additional data on usage of fertilizers and chemicals, farm production expenditures, value of machinery and equipment, value of land and buildings, and farm-related income. A post-stratification estimation procedure is used to produce estimates for the sample items.

Sampling is also used to represent nonresponding farms. Cases that do not respond to the census are divided into large and small depending on their estimated size which is based on expected acres and/or expected value of sales at mail out. The large case nonrespondents number about 100,000. These

cases are followed up on a 100 percent basis using CATI. Large cases not responding to CATI are then classified as farms or nonfarms using secondary source data from the National Agricultural Statistics Service (NASS), the 1987 Census of Agriculture, the Agricultural Stabilization and Conservation Service, etc.

A 100 percent telephone follow-up of small nonrespondents would be prohibitively expensive. For these, we first select a sample of small nonrespondent cases in each county with an overall response rate less than 75 percent. These cases are contacted by CATI in the same way the large nonrespondents are. CATI calls are conducted until a 75 percent response rate is obtained for each county.

Even after the data for the above low response county operations are collected, there are still about a half million small nonrespondents. To estimate how many of these small cases are farms, a sample of about 20,000 nonrespondents, stratified by state and estimated size, is selected. Each of the selected cases is contacted by CATI, or failing that, by certified mail, to ascertain whether it is a farm or not. From this sample, we estimate the proportion of small census nonrespondents that are farms for each county. Weights are assigned to responding farms to adjust for the proportion of nonrespondents that are farms. As information obtained from the respondent is generally believed to be more accurate than imputed data, the percent of the published data that is respondent supplied gives a measure of data quality.

3.5 Data Collection, Processing and Review - Addresses on the mail list are sent up to five follow-up mailings if they fail to respond. The follow-ups consist of a thank you/reminder card, three form follow-ups and a letter follow-up. The follow-ups are sent at approximately 4 week intervals.

Completed questionnaires are returned to the Census Bureau's Data Preparation Division in Jeffersonville, IN. The processing of the report forms includes clerical screening for farm activity, computerized check-in, keying and transmittal of completed report forms, computerized editing of inconsistent and missing data, review and correction of individual records referred from the computer edit, review and correction of tabulated data, and electronic data processing. During census processing, many tables are produced of return rates, response rates, Undeliverable As Addressed (UAA) rates and other processing code rates. The tables provide control data for all geographic levels, for important survey variables, and for significant subpopulations

of the mail universe. These tables provide immediate and frequent information on the status of report forms at all geographic levels. The complex edit and imputation system is designed to ensure reasonable relationships between and among data items, values for various sizes of farms, and combinations of commodities; to ensure necessary consistencies are present (there are more than 70 distinct consistency requirements); and to ensure that geographic, legal, and physical constraints are met. The system performs these and similar functions for nine hundred sample items and eight hundred and fifty nonsample items. A duplicate search operation is conducted during data review in all states by sorting all records alphabetically within counties and matching on telephone number and important data variables.

Quality Assurance/Quality Control activities are conducted for the following census processing activities: mail package preparation, check-in, remove contents and sort, interactive incoming and outgoing telephone calls, correspondence, special case processing, duplicate processing, data entry, edit referral processing, microfilming, coverage evaluation processing, and nonresponse processing.

4. How is Quality Measured for the Agriculture Census?

Several statistical and descriptive measures are calculated to evaluate quality in the agriculture census. The measures are designed to measure sampling and nonsampling error. Among them are response rates, UAA rates, coverage error rates, and other descriptive statistics to evaluate the usefulness of each mail list source, the accuracy of size codes, etc. Many of the measures and a description of the statistical methodology are published in the appendix of the census of agriculture publication. In addition, evaluations are conducted on census processes such as mail list development.

4.1 Sampling Error - Estimates of variability are produced for about 200 of the published census data items. Generalized variance estimates are provided to estimate variances for the remaining items. The variability of the estimates represent the error in the estimates due to nonresponse and sample estimation.

4.2 Nonsampling Error - Return rates, UAA rates, response rates, imputation rates, QC and QA error rates, and evaluation results are all measures used in the agriculture census to help control and measure nonsampling error. Census return rates are the quotient of all receipts (including forms returned by the post office--UAAs) divided by the total number of addresses on the mail list. The return rates for

the 1982 and 1987 censuses were 85.4 percent and 86.7 percent, respectively. The UAAs are essentially nonfarms--addressee is deceased, name or address has changed, or another situation is indicated. In 1987 and 1992, 3.6 and 5.6 percent of the census returns were UAAs. Removing UAAs from receipts and from the total mail list yields response rates.

Response rate is a standard measure of the effectiveness of the data collection in eliciting response. Response rates were 87.1 percent, 87.3 percent, 85.1 percent, 86.2 percent, and 84.5 percent for the 1974, 1978, 1982, 1987, and 1992 censuses respectively. On this basis, the response rate appears to be declining since the 1978 census.

The percent of farms imputed in the census count is 12.6 and 14.3 percent for the two most recent censuses. The proportion of data imputed for other major data items--land in farms, harvested cropland, and value of agricultural products sold--has been consistently lower than that for the farm count.

4.3 Evaluations

Mail List Model - An evaluation of the 1987 and 1992 classification tree analysis (Schmehl, 1990 and Ash, Kraus, Peterson, 1995) included a sample survey of records dropped from the census list, belonging to model groups whose proportion of farms was expected to be 11.7 percent or less for 1987 and 18.8 percent or less for 1992. Approximately 14.6 percent of the 1987 records and 25.9 percent of the 1992 records represented farm operations (46.4 and 82.0 percent survey response, respectively). Approximately 59,378 of the 229,180 records dropped from the 1992 census represented farm operations. From a separate coverage evaluation program, an estimated 254,600 farms were not on the final census list. Using the two estimates, about 23 percent of the farms not on the mail list (59,378 of the 254,600) were on the list before probable nonfarms were dropped. The remaining 77 percent were either on the list of nonfarm addresses which were dropped or not on the list at all. This evaluation and others indicated that the analysis accomplished the objective of identifying the groups of records with questionable farm status and provides a reasonably good estimate of the impact of reducing the size of the final mail list on census coverage.

Mail List Development - Following the 1987 and 1992 censuses, tabulations of the mail list by address sources revealed the contribution of each source. The percent of enumerated census farm records appearing only on one source was 14.9 for IRS, 3.9 for the previous census, and 3.3 for NASS in 1987. The comparable 1992 percents are 13.8 for IRS, 3.3

for the previous census, and 2.0 for NASS in 1992. The number of farms appearing on only one source expressed as a percent of all respondent records for that source was much higher for the IRS list (10.8 and 10.4 for 1987 and 1992, respectively) and the special list (8.7 and 3.6 for 1987 and 1992, respectively) than for other source lists (3.9 and 3.4, respectively for the previous censuses; 4.1 and 2.6, respectively for the 1987 and 1992 NASS list).

Coverage - The Census Bureau has conducted formal coverage evaluation programs for each census of agriculture since 1945. The program measures the accuracy and completeness of farm counts and selected data items and seeks to identify situations that lead to coverage error and to reveal data deficiencies and problems associated with census processes. These errors arise from incompleteness of the census mail list, the complexity of some farm organizational structures, continual changes in the activity status of farms, difficulty in communicating concepts on the questionnaire to respondents, etc.

The evaluation is conducted using an independent sample selected from the mail list to measure classification and list duplication error and the NASS June Agricultural Survey (JAS) to estimate the number of farms not on the census list. Non-farms classified as farms and duplicate operations contribute to overcounted farms in the census. The total number of 1987 estimated overcounted farms was 135,600 (6.0 percent). Farms classified as nonfarms and farms missed on the mail list contribute to undercounted farms in the census. The total number of 1987 estimated undercounted farms was 296,933 (13.1 percent). The net coverage error was an undercount of 161,333 farms (7.2 percent) for the 1987 census. Historical coverage estimates show that net farm coverage of actual farms has ranged from 85.0 to 92.8 percent for all censuses except 1978 when using both a list and area frame resulted in coverage of 96.6 percent. This methodology substantially improved the state and U.S. level coverage of the census, particularly for farms with sales of less than \$2,500 where the census enumeration is least complete. Budget constraints have not permitted the use of multiple frame methodology in subsequent censuses. An estimated 98.6 percent of the land in farms, 92.3 percent of the crop farms, and 87 percent of the livestock farms were on the 1987 census list. Complete 1992 coverage evaluation data is not available at this time.

Nonresponse Survey Study - Another indication of the accuracy of 1987 census address information was obtained from results of an intensive follow-up to the Nonresponse Survey. This survey is conducted prior

to completion of the census data collection for each state to estimate the percent of farms among the nonrespondents. The special study examined a sample of 1,263 nonrespondents to the survey. In early 1989, certified mailings, and telephone and personal enumeration contacts were made to these cases to obtain further information about the nonrespondents. A total of 30.2 percent of the sample cases were UAAs. Of the 31.2 percent for which personal follow-up was attempted, over half had address changes before or during the census. Either the corrected address information had not been received from the Post Office as requested or the Census Bureau had not successfully incorporated the address change into its mail list.

4.4 Quality Control (QC) Measures - Several types of QC procedures are used in the census of agriculture. These include monitoring, acceptance sampling (inspection used to identify and correct errors), process control (monitoring the process on an ongoing basis with adjustment when necessary, using QC results to improve the process), and a combination of acceptance sampling/process control. Both independent and dependent verification is used. Feedback includes control charts, bar charts (type of errors) and verification summary data.

5. What Continuous Improvements are Evident in the Census of Agriculture?

Notable changes in list sources and their content were made for the last two censuses. For these, the Census Bureau used the farm operation list of the NASS for all 50 states as contrasted with the 31 states available for use in the 1982 census list compilation. The NASS list, provided to the Census Bureau prior to linkage operations, enables the census list to include new records and updates to existing records as of April of the census year.

In 1987, the Census Bureau stopped using the Agriculture Stabilization and Conservation Service (ASCS) records directly. This decision was based on the expected inclusion of valid ASCS farms into the NASS lists, the ASCS list size, and the reliability of the ASCS name and addresses. When the ASCS records were used as a census source many ASCS records did not match other source records and required screening. Many of the addressees were determined not to be farm operations for the census. Of the 1982 census farm respondents, only 1.4 percent came uniquely from the ASCS list.

For the 1992 census, refinements were made to the classification analysis model, including the use of standardized computer software (CART) and the use of unique state (rather than regional) model groups.

For the 1992 census, the Census Bureau introduced a probability based linkage system (variant of the Fellegi-Sunter procedure) for name and address matching during mail list development. This system permits the user to specify the degree of certainty (threshold values) desired for matched records. The percent of duplicates detected increased eight percent between 1987 and 1992 for the first linkage phase; one percent in the second phase of linkage. The new software also reduced the amount of clerical review that was required. The threshold values gave us more confidence in the probability of records in a pair being duplicates, so the threshold values were set such that we had an 18.5 percent reduction (in records reviewed by clerks) from 1987 to 1992.

Prior to 1992, the large nonresponse telephone follow-up operation and the nonresponse survey were conducted by clerks in the data preparation unit. In 1992, the follow-up was conducted by CATI.

Research conducted prior to the 1987 and 1992 censuses led to improvements in almost every statistical method used in the census. For example, for the 1987 census the number of initial post-strata used for sample estimation was changed from 128 to 32 initial post strata based on research that revealed that 32 post-strata design yielded a smaller mean square error. Noninteger sample weighting was introduced to insure detail level data added to the totals; a statistical based collapsing procedure for combining strata with insufficient sample cases was used; a raking ratio estimator was introduced for sample estimation to smooth the variation in sample weights; SAS was used to generate models for computing generalized variance estimates; different assumptions and strata were employed for computing nonresponse survey estimates; and classification analysis was used for the first time in MLD. New and refined disclosure avoidance methodology was introduced for the first time in the 1992 census. The coverage evaluation program for the 1987 census of agriculture was enhanced to provide estimates of farms not on the census mail list at the state level as well as more reliable estimates of incorrectly classified operations and duplicate operations on the census mail list. For the 1992 census, the data processing and review for the coverage evaluation data was automated to reduce data processing costs and time and more accurate variance estimators were used.

The Census Bureau is moving further away from using QC to inspect quality into the process and place more emphasis on building quality into the process. This will mean less reliance on inspection to

detect and correct errors and more emphasis on preventing errors from entering the process. To do this, we must reduce the need for inspection, through better equipment/materials, better training, and more employee empowerment.

The census report forms are continually evolving. In 1987, a short form with a section to screen for agriculture activity and a reminder/thank you postcard were used for the first time. In 1992, the style of the report form was changed from a foldout to a booklet form and the short form was dropped. The screener for the short form was placed on the regular census report form. For the 1997 census, a user friendly version of the report form is being tested to ease the respondent burden.

The Census Bureau used several new procedures to help identify duplicates during data review and processing of the 1992 census. To facilitate self-identification of duplicate report forms, the Census Bureau provided instructions to the respondent on the form and information sheet. A statement was added to the census envelope as a reminder to return all duplicate report forms in the envelope.

The Census Bureau is on the leading edge with its use of computer automation in the processing and review of agricultural census data. The automation for census processes spread quickly to all processes such as the automated review of press releases, the implementation of disclosure avoidance, the tabulation of data and the coverage evaluation program. The automation reduced the number of staff and time needed to process and review the data.

6. Quality in Future Censuses of Agriculture?

Providers and users of agricultural data, along with computer technology have continued to change over the years. The Census Bureau has implemented innovations into the Census of Agriculture to accommodate many of these changes. There is still ample opportunity for improving the way the census is conducted. Some things being considered for future censuses include development of methods:

- * for changing organizational and operational structure of large agricultural operations.
- * to provide more economic and demographic data on rural America to meet the needs of rural economic planners and developers.
- * for additional automation in census processing to provide greater efficiency, cost savings and better quality.
- * to further reduce the mail list size and improve the coverage of the farm population.
- * to provide more efficient statistical designs and

more precision.

Conducting a high quality census of agriculture is a monumental task, but one that the Census Bureau has committed itself to for many years. Building quality control and quality assurance into each step yields a quality approach that allows changes, innovations and improvements to be successfully incorporated in the census. Given the ever changing nature of providers and users of agriculture data and computer technology, the Bureau will continue to introduce innovations to maintain and improve the quality of agriculture statistics.

7. Acknowledgements

The authors wish to acknowledge the assistance of a number of colleagues for their help in preparing this paper. In particular, we wish to thank Tommy Gaulden and Anthony Williams, Census Bureau reviewers, for their helpful comments.

8. References

1987 Census of Agriculture Coverage Evaluation Report, Volume 2, Part 2.

1992 Census of Agriculture Geographic Area Series Report, Volume 1.

Ash, Stephen, Melinda Kraus, Anne Peterson (1995), "Evaluation of Classification and Regression Tree Generated Model Groups for the 1992 Census of Agriculture," *1995 Proceedings of the Section on Survey Research Methods, ASA*.

Clark, Cynthia Z.F., E. Ann Vacca (1993), "Ensuring Quality in U.S. Agricultural List Frames," *Proceedings of the International Conference on Establishment Surveys*".

Fellegi, Ivan P. and Alan B. Sunter (1969), "A Theory for Record Linkage", *Journal of American Statistical Association*, pp. 1183-1210, December 1969.

Gaulden, Tommy W. (1990), "Development of the 1987 Census of Agriculture Mail List," U.S. Census Bureau internal report.

Lewis, Paul J., Glenn S. Wolfgang, E. Ann Vacca, "1992 Census of Agriculture Coverage Evaluation Estimation", *1994 Proceedings of the Section of Survey Research Methods, ASA*.

Schmehl, Richard, Magdalena Ramos (1990), "Evaluation of Classification Tree Methodology for Census Mail List Development," *1990 Proceedings of the Section on Survey Research Methods, ASA*.