

CONTENTS

Chapter 10. Coverage Evaluation and Research

	Page
1987 Coverage Evaluation and Research	104
Introduction	104
Background and objectives	104
General procedures	104
Sample Survey Designs and Methodologies	104
1987 June Enumerative Survey (JES)	104
1987 Classification Error Study (CES)	104
Data collection	105
Processing	105
JES file processing	105
CES processing	105
Estimation Procedures	105
Results and Publication	105
1987 Advertising Awareness and Response Behavior Study	106
Introduction	106
Scope and objectives	106
Sample design and selection	106
Questionnaire content	106
Telephone Operations	106
Results	107
Classification Tree Methodology Evaluation	107
General Information	107
The Evaluation Methodology	107
General procedures	107
Measures of association	107
Evaluation by farm frequencies	107
Survey data	108
Evaluation Results	108
Disclosure Avoidance Research Study	108
General Information	108
Research	108

1987 COVERAGE EVALUATION

Introduction

Background and objectives—The Census Bureau first carried out a coverage evaluation of the agriculture census in 1945, and first released the results of such a study as part of the 1950 census publication program. Since then, the Bureau has routinely evaluated each agriculture census for the accuracy and completeness of the published farm count, and for coverage of selected data items, such as land in farms, total value of agricultural products sold, and so on. The basic methodology used has remained largely unchanged—an area and list sample survey—although methods and sample designs have been refined and improved with each census.

The primary objectives of the 1987 Census of Agriculture coverage evaluation program were to provide—

- State estimates of the number of farms not on the mail list
- Regional estimates of the number of farm operators incorrectly classified and duplicate farms
- Divisional estimates of selected agricultural characteristics of undercounted farms

The 1987 coverage evaluation program estimated that 242,853 farms were not on the census mail list; 54,080 farms were omitted from the census tabulations due to incorrect classification as nonfarms; 72,310 nonfarms were counted as farms; and 63,290 farms returned more than one census report form.

General procedures—The 1987 coverage evaluation program used U.S. Department of Agriculture's (USDA's) National Agricultural Statistics Service's (NASS's) 1987 June Enumerative Survey (JES) and the 1987 Classification Error Survey (CES—part of the 1987 agriculture census program) for developing its coverage estimates. The JES is an annual national area sample survey designed to measure planted acreage of crops and numbers of livestock by State. NASS conducted the JES, and the Census Bureau used the 1987 JES data to estimate the number and characteristics of farms not on the census mail list. The 1987 CES data were used to estimate the number of misclassified farms (i.e., farms incorrectly classified as nonfarms, and nonfarms as farms) and duplicate farms in the census.

The JES was a field canvass of selected area segments, and the area segment records then were matched to the census mail list. Matched and nonmatched addresses were mailed census report forms, the matched ones as part of the regular census and the nonmatched cases as additional census mailings. The resulting data were used to compute the "not on the mail list" estimates. The 1987 CES used mail enumeration, with followup by telephone to

nonrespondents. The survey farm status (i.e., farm or nonfarm) was compared to the census farm status to determine whether a given sample farm had been correctly classified in the census. The results of the survey operations were processed, tabulated, analyzed, and published.

Sample Survey Designs and Methodologies

1987 June Enumerative Survey (JES)—The USDA's National Agriculture Statistics Service (NASS) conducted the June Enumerative Survey (JES) annually to measure planted acreage of crops and numbers of livestock. The NASS area sample frame consisted of land parcels of predetermined size, with easily identifiable boundaries. All land in each State was stratified based on land use, such as intense cultivation, urban areas, range land, and so on. Cultivated land was divided into several strata based on the degree of cultivation throughout the State. Primary sampling units (PSU's) were land parcels selected within each land use stratum. Each PSU was divided into several ultimate sampling units, or segments. PSU size varied, but typically included six to eight segments, each with an average of three farms.

The JES used a two-stage probability area sample of farm operations. In the first stage, the survey selected PSU's with probability of selection proportional to the number of segments within each PSU of each land-use stratum. In the second phase of the JES, a segment was selected with equal probability from each PSU for enumeration. Each distinct farm operation within a segment was defined as a tract. Any land area within a segment that contained one or more occupied dwellings was designated a residential tract. To increase the reliability of the estimates of farms not on the census mail list, the NASS retained in the sample for the 1987 JES the 20 percent of the agricultural-urban segments that normally would have rotated off the sample in 1987. At the Census Bureau's request, NASS also instituted more stringent screening procedures for residential tracts, requiring its field interview staff to list all houses in each segment and to inquire whether any household members or neighbors were involved in agricultural operations.

1987 Classification Error Study (CES)—The 1987 CES was designed to measure the number of farms on the census mail list that were incorrectly classified or were erroneously duplicated. The Bureau staff drew a systematic sample, stratified by geographic region, from the initial census mail list (4.1 million addresses), excluding addresses (1) in Alaska and Hawaii, (2) with \$500,000 or more expected annual total value of sales of agricultural products, (3) representing abnormal or multiunit operations, and (4) representing JES nonmatch records. The CES sampling rate varied by census geographic region (see ch. 5 for a description of the census divisions and regions), as follows:

Region	Sample rate
Northeast	1 in 71
Midwest	1 in 500
South	1 in 176
West	1 in 227

Approximately 18,500 names and addresses were selected, with sufficient numbers in each region to provide acceptable regional-level error estimates. At specified cutoff dates, the Bureau obtained the report form check-in status for the CES sample addresses, using the CES evaluation code set in the census data base at the time of the initial sample selection.

Data collection—NASS collected the data for the 1987 JES by field canvass of each segment in June 1987. The Census Bureau received the JES data files containing names, addresses, and agricultural data on all area-segment residents involved in any agricultural activity. It matched this file to the census mail list development file. All JES records not on the census mail list were assigned a special processing code for identification and were added to the census mail file.

The 1987 CES was a mail and telephone enumeration. Form 87-A90, Evaluation of the 1987 Census of Agriculture, report forms was mailed to CES sample addresses in two waves (according to the date of response), the first in March, and the second in July 1988. The mailouts went to 15,331 sample survey cases that had responded to the census, with a reminder card followup 14 days after the initial mailout, and a second report form 14 days after the card. Six weeks after the first CES mailing for each wave, the Bureau referred nonrespondent cases (approximately 4,200) to the Data Preparation Division's (DPD's) telephone unit in Jeffersonville, IN, which carried out a telephone followup between the second week of June and the end of September 1988.

Processing

JES file processing—Processing clerks identified and reviewed area segment survey records using both census and JES data, such as the type of crops, livestock, total value of sales of agricultural products, and land use, to determine whether a JES case that did not respond to the census qualified as a farm under the census definition, and to resolve cases in which JES and census farm status differed. The clerks added a coverage classification code to each record to specify the farm status (i.e., farm or nonfarm) and the final match status of the survey record to the census mail list (match or nonmatch). Both the match and nonmatch records were retained in the sample for deriving estimates of farms not on the mail list.

The Jeffersonville staff keyed the JES data to a computer file, which then was edited using the interactive minicomputer system, and reviewed for consistency and accuracy prior to tabulation. The Bureau used the JES

nonmatch records to estimate the total number and selected characteristics of farms not on the census mail list for States, divisions, regions, and the United States.

CES processing—The Jeffersonville clerical staff edited the CES sample forms for consistency and accuracy, then carried out a separate technical review to classify agricultural operations as farms or nonfarms. The processing clerks compared the CES data and farm status to the census data and farm status, and assigned coverage classification codes that identified farm operations and match/nonmatch status between each record's CES farm status and census farm status. The data then were keyed to tape and transmitted to the Suitland facility for computer editing and tabulation.

Estimation Procedures

The Bureau used the final data file from the CES in conjunction with JES data to produce regional and U.S. net coverage-error and classification-error estimates for the census itself. The estimated true total number of farms in the United States is the census published farm count minus the number of overcounted farms plus the number of undercounted farms. The coverage evaluation provided estimates of (1) undercounted farms (farms not on the mail list and farms incorrectly classified as nonfarms); and (2) overcounted farms (nonfarms incorrectly classified as farms, and duplicates). In these estimates, the Bureau used a model that consolidated estimates from two enumerations—in this case, NASS's 1987 JES and the 1988 CES—to estimate a true total. The model assumed that (1) both the census and the JES attempted to enumerate accurately the complete universe of farms, and that farms reported on either source list were true farms; (2) the event of being included in the census was independent of the event of being included in the survey; (3) the probability of being missed by either the census or the survey was the same for all farms within a given size category; and (4) every farm in the complete universe of farms had, independently of every other farm, the same chance of being listed in the census and, independently again, of being listed in the JES. (For details of the estimation methodology, see the *1987 Census of Agriculture, Volume 2, Subject Series, Part 2, Coverage Evaluation*.)

The statistical model and the data from the JES and CES then were used to estimate the undercount and the overcount. The coverage evaluation produced State-level estimates of the number and characteristics of farms not on the mail list, and region-level classification-error estimates for the number of incorrectly classified and duplicate farms.

Results and Publication

The 1987 coverage evaluation program estimated the net farm coverage undercount (including classification error) to be 7.2 percent (plus or minus 0.5 percent), with a

resulting census coverage of 92.8 percent. Estimated total undercount was 13.2 percent, with an overcount of 6.0 percent (both plus or minus 0.5 percent). The undercount and the overcount both were highest for small farms—the estimated undercount was 32.3 percent for farms with annual sales of \$2,500 or less, and 27.9 percent for farms with less than 50 acres—while livestock specialty operations tended to have a higher undercount than other farms, and general crops operations a slightly higher overcount than other operations.

The State-level estimates of the number and characteristics of farms not on the mail list appeared in the Volume 1, *Geographic Area Series* report for each State. The State-level estimates for farms not on the mail list, as well as the classification error estimates, were released in a printed report, the *1987 Census of Agriculture, Volume 2, Subject Series, Part 2, Coverage Evaluation*. The publication included text describing the coverage evaluation program; charts and tables showing estimates for the United States and regions of census farm coverage; farms by selected characteristics and components of coverage; selected items for undercounted farms; and both land in farms and value of agricultural products sold, by sales group and components of coverage.

1987 ADVERTISING AWARENESS AND RESPONSE BEHAVIOR STUDY

Introduction

Scope and objectives—The success or failure of the agriculture, or any other, census, depends entirely on the degree of cooperation and response from the individual operators. Consequently, the Census Bureau was interested in identifying the factors that affect census response. It planned and carried out the 1987 Census of Agriculture Advertising Awareness and Response Behavior Survey (ARBS) to help identify the various events, opinions, knowledge, or other factors, that influenced census response.

The 1987 ARBS was a telephone survey of a sample of agricultural operations drawn from the 1987 Census of Agriculture mail list. Staff contacted operators in the sample by telephone and interviewed them about their awareness of the census, exposure to the agriculture public information campaign, opinions on Federal and local government use of census data, ability to complete the census form, intentions to respond, and so on.

The Bureau used the results of the survey and analysis for internal planning purposes and did not publish a formal report.

Sample design and selection—The ARBS sample consisted of a control and four sample panels, selected during the final stage of mail list development. The sample design excluded all addresses of operations with expected annual sales of agricultural products of \$500,000 or more, multiunits, specialty operations (i.e., known horticultural or

other specialty operations), abnormal, and all addresses in Alaska and Hawaii. These places were excluded because the Bureau employed different data collection procedures for them. The Agriculture Division staff selected national samples of approximately 1,900 addresses each for panels 1, 3, and 4, and 7,600 addresses for a regionally stratified panel 2. The sample size was determined by the requirement that the research staff be able to compare respondents and nonrespondents within each panel, between panels, and between regions for panel 2. Each panel was designed to examine the effectiveness of different publicity campaigns and followup procedures by different data collection activities throughout the census.

Questionnaire content—The ARBS report forms were designed to be completed in a 10-minute telephone interview. Each form contained three parts; parts A and C were essentially identical, while part B differed slightly for each panel, to collect specific information to assess the respondent's awareness of the census at four specific points in the data collection process.

Part A requested information to (1) confirm the identity of the respondent and that he or she operated a farm, (2) assess the interviewee's awareness of the agriculture census from various media, and (3) determine whether the interviewee knew census response was mandatory. Part B asked for information on (1) the respondent's knowledge of when the next census would be conducted (all panels) and of the due date for census response (panels 2, 3, and 4); (2) whether the respondent had received the initial and subsequent mailings (panels 2, 3, and 4); (3) the interviewee's response to the mailings (i.e., whether he or she opened the envelope, read the letter, used the information sheet, filled out or started to fill out the report form, and so on) (panels 2, 3, and 4); and (4) the respondent's impression of the length of the census form and the time required to fill it out. Part C of each form was designed to obtain the respondent's views and beliefs about the agriculture census program and the uses of the census data, and to determine his or her level of education and age range.

Telephone Operations

The Bureau initially used its computer assisted telephone interviewing (CATI) facility in Hagerstown, MD, for panel-1 calls. The much larger sample size of panel 2, combined with other commitments for the CATI operation, led the agency to shift the telephone operation to the telephone unit at the DPD office in Jeffersonville, IN.

After sample selection, the names and addresses for each panel were referred to telephone number research units (first at Hagerstown, for panel 1, and at Jeffersonville for the other three panels). The CATI staff telephoned panel-1 cases during the first 3 weeks of December 1987, before the mailout for the agriculture census. Panel-2 telephone calls were made by Jeffersonville staff through January 1988, after the initial census mailing but before the reminder/thank you card was received; panel-3 cases

were called in March, after the reminder/thank you card and first followup packages had been mailed. The telephone interviewers contacted panel-4 cases in the last week of May, by which time the Bureau had mailed the initial census report forms, the reminder/thank you card, and five nonrespondent followups.

Results

The Agriculture Division's Research and Methods Branch designed the ARBS to measure the effectiveness of the public awareness campaign and followup procedures, and to enable analysts to make comparisons between the samples and between census respondents and nonrespondents. After completing the telephone interview operation in May 1988, Agriculture Division tabulated and analyzed the data from respondents.

Individuals most frequently reported that the source of media exposure that informed them of the census were agricultural organization meetings and other sources, such as previous censuses, posters in stores or post offices, flyers received in the mail, co-workers, colleagues, and so on. The exposure to the census campaign also appeared to have an effect on the survey respondents' knowledge, attitude, and opinions. Higher knowledge, attitude, and opinion scores were consistently observed from sample respondents who had been exposed to the publicity materials, than from those who had not received any information about the census. This was true also in comparing census respondents and nonrespondents; respondents had better knowledge of the census, and more favorable attitudes and opinions toward the census, than did nonrespondents.

The study also showed that the later sample (panel 4) and the nonrespondent sample appeared to (1) need more time to complete the form, (2) feel that the census report form was too long, and (3) not know the census response due date.

The ARBS data suggested that the public awareness campaign and followup activities—when considered as one continuing activity throughout the census period—appeared to have a positive effect on improving the public's knowledge and opinion about the agriculture census. Evaluation of the ARBS data also indicated certain aspects of the publicity and followup procedures that needed to be improved for future censuses.

CLASSIFICATION TREE METHODOLOGY EVALUATION

General Information

The Census Bureau used a binary classification tree procedure (see ch. 3) for statistical modeling to group 1987 preliminary mail list addresses according to their likelihood of being farms. The addresses identified as less likely to be

farms were excluded from the 1987 mailout. The Agriculture Division staff evaluated the technique's effectiveness. Data for the evaluation were drawn from three sources: (1) results from the use of the methodology (i.e., the proportion of farms in the model group), (2) observed responses to the 1987 census, and (3) observed responses to an independent survey of addresses removed from the census mail list.

The Evaluation Methodology

General procedures—The classification tree methodology evaluation, carried out following census data collection and processing, consisted of comparing *expected* farm proportions and frequencies computed during the application of the methodology with the *observed* farm proportions and frequencies for both agriculture census records and the specific records excluded from the census mail list by the application. The evaluation was done in three phases, using (1) measures of statistical association, (2) expected and observed farm frequencies, and (3) an evaluation of data obtained from the survey of addresses removed from the mail list.

Measures of association—The first phase of the evaluation involved partitioning the data file by ranges of expected farm proportions and by observed farm/nonfarm status, and the creation of two row-and-column contingency tables for use in analyzing the accuracy of the classification tree procedure. For both tables, the Bureau used observed farm status in the 1987 census to create row classes, while the column categories were based on the expected proportion of farms. The measures of association enabled the Bureau to determine if the observed census farm status was correlated with the expected farm proportion. The agency used the likelihood of an address being mailed one of the form types (short or regular) to establish categories—one for groups of addresses considered to have more than 11.7 percent, but less than 43.22 percent farms, and the second for the group of addresses considered likely to have over 43.22 percent farms. During the statistical modeling used in the compilation of the census mail list, groups of addresses considered to include 11.7 percent or fewer farms were deleted from the list. Groups of addresses believed to include more than 11.7 percent, but less than 43.22 percent, farms were added to the mail list for the short form, while groups believed to include more than 43.22 percent farms were included in the regular census report form mailings.

Evaluation by farm frequencies—In the second phase, the data used for the evaluation were the expected and observed farm frequency counts from 1,839 of the 2,184 statistical model groups created by the classification tree methodology in its application to the census mail file. (There were no observed farms for 345 of the classification tree model groups, so those groups had to be dropped from the evaluation.) To be able to compare the frequency

counts, the expected frequency counts were calculated by multiplying the observed number of 1987 respondents in a model group by the group's expected farm proportion. By comparing the expected and observed farm frequencies for the model groups, the evaluation staff could assess any differences or similarities in expected and observed farm frequency distribution, which would indicate if the expected and observed farm proportions differed.

Survey data—The third phase of the program employed data obtained from the Model Drop Survey, conducted in dropped the summer of 1988. Approximately 5,300 addresses that had been deleted (dropped) from the census mail list as a result of the classification tree methodology were canvassed by mail to determine their farm status. The records selected for the survey, and responses obtained, were assigned to five strata based on

1. Source list (strata A-C)
2. Cases identified by the classification tree methodology but not included in strata A-C (stratum D)
3. Those subjectively excluded from the mail list by Agriculture Division personnel because they believed these records were unlikely to represent farms (stratum E).

For evaluation purposes, the records in each stratum also were assigned to three categories based on expected proportion of farms assigned to them by the classification tree methodology. While the survey obtained responses from only about half of the addresses mailed questionnaires, the staff had sufficient data collected to analyze response and inflate the results statistically for use in the evaluation. The objective was to determine whether a comparison of the observed survey-wide farm proportion to the minimum census inclusion proportion (0.117) dictated that the survey cases (and the universe they represented), or certain categories of the survey cases (and their universe(s)), should have been included in the 1987 census. These comparisons established a reference point from which to conduct further analyses, if needed.

Evaluation Results

The general conclusion of the evaluation was that the classification tree methodology was successful in selecting which addresses should be included in the final 1987 agriculture mail list, but that the accuracy of the procedure could be further improved. The results of the various measures of association indicated that the classification

methodology performed fairly well and accurately determined the proportion of farms in each of the model groups, while the Model Drop Survey results showed that expected farm proportions were significantly lower than the observed proportions for five of the eight strata and model group categories. The Agriculture Division staff also determined that the number of cases composing the final model groups were sometimes too small to estimate farm proportions, and recommended further research in that area.

DISCLOSURE AVOIDANCE RESEARCH STUDY

General Information

The Bureau published agricultural census data in tables, some relatively simple, and others made up of detailed cross tabulations (see ch. 5). The confidentiality rules of the census law required that no data be published that could be used to identify a specific respondent or operation, hence the tabulations had to be reviewed and items that would disclose identities were suppressed (i.e., not shown in the detailed tables), although the data were tabulated for total counts.

The 1987 disclosure procedures involved applying initial suppressions to any value that failed a predefined disclosure rule, and then examining data items involved in the initial suppression for complementary suppression. Initial suppressions were data suppressions required by the disclosure avoidance rules. Complementary suppressions ensured that none of the initially suppressed values, or previously applied complementary suppressions, could be derived by the addition or subtraction of published data. Published farm counts were not considered a disclosure and, thus, were not subject to suppression.

Research

In 1989, the Bureau's Agriculture and Statistical Research Divisions began a research program to study the complementary suppression methodology used in the 1987 census. Researchers examined the possibility of employing a new system of applying complementary suppression that would increase data utility by providing (1) a lesser number of complementary suppressions, (2) a smaller sum of all data values suppressed as complements, (3) a consistent level of protection, and (4) an improved reconciliation program between different cross tabulations and between the master matrix and the cross tabulations. This research was directed at improving the Bureau's suppression methodology, and continued in subsequent years.