

COVERAGE EVALUATION FOR THE 1982 CENSUS OF AGRICULTURE

William C. Davie, Emily Lorenzen, and D. Dean Prochaska, U.S. Bureau of the Census

1. INTRODUCTION

A coverage evaluation program for a census is an important means for assessing the completeness and accuracy of the data. The purpose of this paper is to present the objectives, sample design, methodology, and preliminary results from the 1982 Census of Agriculture Coverage Evaluation Program. The results will not be used to adjust the census data but will provide independent measures of the number and characteristics of farms not on the census mail list, misclassified farms, and overcounted farms. The program also aids in the identification of separate problem areas for future improvements in developing the census mail list and in collecting and processing the data.

1.1 Background

The census of agriculture currently is taken every five years to provide a detailed statistical picture of a vital sector of the Nation's economy. The first census of agriculture was taken in 1840 as part of the decennial census and every ten years until 1925 when it was taken generally on a 5-year cycle. In 1976, the agriculture census reference years were changed to coincide with the other economic censuses. The change in the reference years provided for joint processing operations and more data comparability among the various censuses.

The 1982 Census of Agriculture was the 22nd nationwide census taken by the Bureau of the Census and the fourth with the data collected by mail. The 1978 Census of Agriculture used an area sample to supplement the census mail list. In 1982, an area sample was not conducted for the census.

Over time, the census definition of a farm operation has undergone several changes. Since 1974, a farm has been defined as any place from which \$1,000 or more of agricultural products were sold or potentially could have been sold during the census year. A place not having sufficient sales to qualify as a farm could qualify on potential sales based on the inventory and production of crops and/or livestock.

The initial step in conducting the census was the development of a mail list from various administrative source lists.[1] The primary source lists used were Internal Revenue Service files, U.S. Department of Agriculture files, and files from the previous census. All of the source lists contained units which were indicated as being associated with agriculture but were not necessarily census farms. Since a name for the same farm could appear on more than one source list, a two-phase computerized record linkage operation was performed to identify duplication. The entire record linkage operation and a screening survey reduced the source lists of about 19.0 million names to approximately 3.6 million names for the census mail list. The report forms were mailed in late December 1982 to the names on the census mail list. A reminder card and five follow-up letters were sent to nonrespondents at 3- to 4-week intervals.

Telephone calls were made to nonrespondents who were expected to have large operations (expected sales of \$100,000 or more) or who were located in low response counties. A nonresponse adjustment procedure was used to represent the final nonrespondent farms in the census results.[2] The data were processed through an extensive computer edit and review. The final stage was the tabulation and publication of the census data for all counties, states, regions, and the Nation.

1.2 Coverage Evaluation Objectives

Although the goal of each census is to include all farm units, it cannot realistically be attained. Complexity of farm organizational units, continuing change in operational units, inadequacies of source lists, difficulty in communicating census definitions and concepts, and other factors each can contribute to census error and incompleteness. An evaluation of coverage has been completed for each agriculture census since 1945. Coverage evaluation programs are designed to provide independent measures of completeness and accuracy of farm counts and selected data items for the censuses of agriculture.

The 1982 Census of Agriculture Coverage Evaluation Program was planned to accomplish several objectives. These objectives were:

- (1) Provide U.S. and regional level measures of accuracy of census farm counts and a limited number of other items, such as land in farms and value of agricultural products sold.
- (2) Provide estimates indicating the characteristics of undercounted (missed) farms.
- (3) Provide information on factors associated with census error to identify areas for improvement of coverage and quality in future censuses.

2. SAMPLE DESIGN AND METHODOLOGY

2.1 Sample Design and Data Collection

The 1982 Coverage Evaluation Program consisted of two parts:

- (1) An area segment survey. This survey was designed to measure the number and characteristics of farms operated by persons living in rural areas (areas with population less than 2,500 inhabitants in the 1970 Census of Population and Housing) who were not on the census mail list.
- (2) A classification error study. The study was designed to measure the number and characteristics of both rural and urban (areas with 2,500 or more inhabitants in the 1970 Census of Population and Housing) farms on the census mail list, but overcounted or misclassified.

The census mail list included farms operated by persons residing in both urban and rural areas. The 1978 coverage evaluation program found that less than one percent of all farms were urban farms not on the mail list.[3] Due to budget restraints, no study was done to measure the urban farms not on the mail list in 1982.

2.2 Area Segment Survey

The 1982 Coverage Evaluation Area Segment Survey was a sample of 344 segments selected from the 6,400 segments used in the 1978 Census of Agriculture Area Sample (CAAS). The CAAS was designed to supplement the 1978 census mail list by providing state-level data for farms that were not included on the mail list. Both the 1969 and 1974 coverage evaluation studies had indicated that there was a relatively high rate of mail list undercoverage for small farms (sales less than \$2,500) and, to a lesser degree, midsize farms (sales between \$2,500 and \$40,000). The sample frame for the CAAS was constructed from maps and data from the 1970 Census of Population and Housing. It consisted of rural areas with population less than 2,500 inhabitants. The CAAS was a stratified one-stage cluster sample with stratification by state and by a farm density ratio (the ratio of farm households to total households). Approximately 6,400 segments were selected systematically across the United States excluding Alaska and Hawaii. The sample was allocated to the strata in a way that approximated an optimum allocation. The segments were selected with uniform probabilities within each stratum. However, the sampling rates varied substantially between strata. The estimated average number of farms per segment was ten and varied by strata from none in residential type areas to 12 in intensive farm areas.[4] For the 1982 area segment survey, the 6,400 segments used in the 1978 CAAS were stratified by geographic region: Northeast, Midwest (formerly North Central), South, and West and number of farms identified as undercounted in the CAAS: 0, 1, 2 or 3, 4 or more. Once ordered, a measure of size based on the 1978 CAAS weights was assigned to each segment in a stratum. A sample of specified size was then selected from each stratum with probabilities proportional to the measures of size. Measures of size were used in the selection procedure as a means of providing a sample for the 1982 area segment survey which had approximately the same overall selection probabilities in each of the 16 strata. The use of these measures of size provided an adjustment for the variation of the 1978 segment selection probabilities within each of the 16 strata.

In order to achieve an absolute standard error of 2.0 percent for the estimated proportion of farms not on the mail list at the regional level, it was necessary to select a sample of 344 area segments. Consideration was given to cost, precision, and availability of maps in sample selection. Regional sample sizes were based on an optimum allocation of the sample, with the constraint that at least two segments be allocated to each strata within the region. Table 1 gives the regional distribution of sample segments. The selection probability for each segment chosen for the 1982 area segment survey was the product of two factors: (1) the selection probability for the 1978 area sample, and (2) the conditional probability of selection for the 1982 area segment survey. The final weight assigned to each segment selected was the inverse of the overall selection probability.

Table 1. Estimated Absolute Standard Error of Proportion of Farms not on Mail List by Region

Region	Sample Segments	Estimated Absolute Standard Error (percent)
United States	344	1.2
Northeast	127	2.0
Midwest	32	2.0
South	92	2.0
West	93	2.0

The data collection procedures were uniform across all segments with extensive emphasis on completeness and accuracy. Experienced enumerators were chosen to canvass the segments. The enumerators completed intensive self-study materials prior to receiving their assignments. The enumeration began in February 1983. Enumerators visited each household in the segment and listed the name and address of the reference person (usually the owner or renter) for each household. The reference person was asked a short series of screening questions to determine if any person in the household had any agricultural operations in 1982. A farm was included in a segment if the farm operator lived inside the segment boundaries. A responsible person was asked the questions if the reference person was not available for interviewing.

For those households having agriculture activity, an "Evaluation of the 1982 Census of Agriculture" report form was completed for each agriculture operation in the household. The evaluation report form was an abbreviated version of the census report form. The evaluation form contained questions about alternate farm names and addresses used for the operation; questions on farm size, crops, and livestock; and questions on various operator characteristics. Enumeration was completed by May 1983, and all evaluation report forms were returned to the Washington, D.C. office for processing.

2.3 Classification Error Study

Classification error also contributes to coverage error in the census of agriculture. Coverage evaluation studies for recent censuses have shown that about 3 to 5 percent of all farms on the mail list were misclassified as nonfarms. Another 1 to 2 percent of all farms were incorrectly classified as farms or had more than one report in the census and were therefore overcounted. Classification error results from misinterpretation of census definitions and instructions by respondents or from errors in census processing.

The sample for the classification error study was a multistage sample selected from the census mail list of 3.6 million names and addresses. Addresses in Alaska and Hawaii were excluded because of limited evaluation funds. Farms with expected sales of \$500,000 or more, institutional farms, and a small number of complex organizational units were excluded because all such farm operations received extensive census mail

follow-up, telephone follow-up, and report form review to ensure the accuracy and completeness of their data. The first selection stage was a systematic sample of about 12,000 names and addresses selected from the census mail list with a sampling rate that varied by census geographic region: 1 in 75 in the Northeast, 1 in 500 in the Midwest and South, and 1 in 150 in the West. These rates resulted in approximately equal numbers of names and addresses from each region. The second selection stage was a systematic sample of about 4,700 names and addresses selected from the first stage sample with an overall rate of 2 in 5. With consideration for cost and precision, this sample was of sufficient size for regional estimates of classification error. Table 2 shows the sample size and estimated absolute standard error of the proportion of misclassified and overcounted farms by region.

Region	Sample Size	Estimated Absolute Standard Error (Percent)
United States	4,681	0.4
Northeast	1,094	0.8
Midwest	1,112	0.5
South	1,206	0.7
West	1,269	0.7

The census report forms that were mailed to the 4,700 sample addresses were identified with a special identification symbol on the name and address label of the report form. It was used only to separate the sample report forms for photocopying when they were returned by the respondents. The respondents and processing staff were unaware of the special symbol, and the forms were returned to regular processing after being photocopied. The photocopies of the sample report forms were reviewed and classified as: farms -- 2,700 farms; nonfarms -- 1,400 farms; nonrespondents -- 500 farms; and Post Master Returns (PMR's--undeliverable by post office) -- 100 forms.

The third selection stage was the study sample chosen for reenumeration. It consisted of all nonfarm cases, all PMR's and a subsample of the farm cases. Of the 2,700 farm cases, about 300 cases were systematically sampled with rates that varied by census geographical region: 1 in 5 in the Northeast, Midwest, and South, and 1 in 7 in the West. These rates were used to provide the desired level of accuracy for regional data.

Data for the classification error study were collected primarily by telephone interviews. For the telephone reenumeration, trained interviewers conducted detailed probing interviews using the evaluation report form. If a household could not be contacted by telephone or a telephone number could not be obtained, an evaluation form was sent by mail with a cover letter to explain the purpose of the study. About 13 percent of the

telephone interviews could not be completed. Most often the household could not be contacted by telephone, and in a small number of cases the individuals refused to be interviewed or gave incomplete data. Incomplete cases were assumed to be similar to the interviewed cases and no separate adjustment was made to the data. This would result in a downward bias in the estimates for misclassified and overcounted farms and a small downward bias in the estimated totals.

The evaluation report forms were compared to the census report forms to identify errors. The types of census errors that were detected were: (1) farms misclassified as nonfarms, (2) nonfarms incorrectly classified as farms, (3) more than one report form for the same farm, and (4) farms that were PMR's in the census ("not classified" in the census).

The 500 census nonrespondents in the classification error study were not reenumerated. Non-respondent farms were represented in the census by a statistical adjustment procedure and were not uniquely related to individual census records. Therefore, classification error could not be measured for this group. A separate study of the statistical adjustment procedure is planned as part of the data collection procedures research.

2.4 Processing

The principal processing steps for both the area segment survey and the classification error study were similar. For both studies, the evaluation report forms were reviewed and classified as farm or nonfarm according to the farm definition. The evaluation report forms were then matched to the census mail list using all information obtained from the interviews, including alternate names and/or addresses. When an area segment farm was matched to a farm on the census mail list, the area segment farm was considered to be a "matched farm." Nonmatched area segment farms were considered to be "farms not on the mail list." In the classification error study, misclassified cases were farms misclassified as nonfarms and farms that were PMR's in the census. Overcounted cases were nonfarms incorrectly classified as farms and farms with more than one census report form.

Following the matching, a final review accompanied by telephone follow-up, if needed, was completed. Coverage classification codes were assigned reflecting types of census errors. The data were then keyed, edited, and reviewed for accuracy and consistency. The final steps were tabulation of data, production of estimates, calculation of variances, and publication of results.

3. PRELIMINARY ESTIMATES OF CENSUS COVERAGE

The preliminary estimates of census coverage for the number of farms are based upon the area segment survey and the classification error study. Estimates shown in Table 3 are for numbers of farms only. Measures for value of sales and land in farms are not included in the preliminary data but will be presented in the final published results. Estimates of sampling reliability are provided with the data.

3.1 Estimation Procedure

Keeping in mind that measures of urban farms not on the mail list were not available in 1982, an estimation procedure was chosen that would account for, in part, the absence of this portion of the urban farms. The estimation procedure assumes that urban farms not on the mail list are similar to the remainder of the farm universe. This assumption is supported by data from the 1978 coverage evaluation.

Any total (T) for some characteristic of all farms in the United States can be represented as the census published number for that characteristic (C) plus the undercount for that characteristic (U) minus the overcount for that characteristic (O). In symbols: $T = C + U - O$. In the 1982 Coverage Evaluation Program, the overcount was estimated for the farm count only.

The undercount (U) can be broken into two components, the part of the universe not on the census mail list (M) and the part of the universe for farms on the census mail list that were misclassified as nonfarms (MCF). In symbols: $T = C + M + MCF - O$. The estimates of the overcount (\hat{O}) and the total of some characteristic for farms on the census mail list misclassified as nonfarms (MCF) are unbiased sample estimates from the classification error study. The estimate for the total of some characteristic for farms not on the census mail list (\hat{M}) is based on a coverage error model which is equivalent to a capture/recapture model.[5]

The coverage error model assumes that the census (list A) which is observable and the universe of the area segment survey (list B) which is not observable both attempt to accurately enumerate the complete universe of farms and that farms reported on either list are true farms. The estimated number of farms on both list A and list B is the unbiased sample estimate from the area segment survey (\hat{N}_i). The estimated number of farms on list B but not on list A is the unbiased sample estimate from the area segment survey (\hat{N}_s). The number of farms on list A but not on list B is estimated by $(N_c - \hat{N}_i)$ where N_c is the published number of farms in the census.

Finally, the number of farms not on list A and not on list B is estimated by:

$$\hat{N}_r = \frac{\hat{N}_s (N_c - \hat{N}_i)}{\hat{N}_i}$$

Thus the estimate of the total number of farms not on the census mail list is $\hat{N}_s + \hat{N}_r$.

The component of the undercount of some characteristic for farms not on the census mail list (M) was estimated by:

$$\begin{aligned} \hat{M} &= (\hat{N}_s + \hat{N}_r) \left(\hat{S} / \hat{N}_s \right) \\ &= \frac{N_c}{\hat{N}_i} \hat{S} \end{aligned}$$

where \hat{S} is the unbiased sample estimate of the total of some characteristic for farms on list B but not on list A. As an estimate of the total number of farms not on the census mail list \hat{M} becomes $(N_c / \hat{N}_i) \hat{N}_s$. Rewriting the total T for some characteristic in terms of the estimated

components obtains:

$$\hat{T} = C + \hat{S} (N_c / \hat{N}_i) + \hat{MCF} - \hat{O}$$

\hat{S} and \hat{N}_i are regional level estimates derived from the sum of the estimates for farms with sales less than \$2,500 and farms with sales \$2,500 or more. The estimates of U.S. totals are the sum of the regional level estimates.

The estimates of the proportions for some characteristic of farms not on the census mail list, misclassified farms and overcounted farms are in the form:

$$\text{Farms not on the census mail list (percent)} = \frac{\hat{M}}{\hat{T}} (100).$$

$$\text{Misclassified farms (percent)} = \frac{\hat{MCF}}{\hat{T}} (100).$$

$$\text{Overcounted farms (percent)} = \frac{\hat{O}}{\hat{T}} (100).$$

3.2 Preliminary Results

The preliminary estimates of census coverage for number of farms are presented in Table 3 by region. Columns 1 and 2 show the estimated percent and the relative standard error for farms not on the mail list derived from the area segment survey. Columns 3 through 6 show the estimated percents and the relative standard errors for misclassified and overcounted farms derived from the classification error study. The farms not on the mail list (10.6 percent), plus the misclassified farms (3.1 percent), minus the overcounted farms (4.6 percent), are equal to the estimated net undercounted farms (9.1 percent). Subtracting the estimated net undercounted farms (9.1 percent) from 100.0 percent gives the estimated percent of all farms that were included in the census (90.9 percent) in the conterminous United States.

Coverage for larger farms (sales of \$2,500 or more) was more complete with 99.5 percent included in the census compared to 71.5 percent of smaller farms (sales of less than \$2,500) included in the census. The primary reason for the more complete coverage of larger farms was that larger farms were more likely to be included in the administrative source lists. In addition, operators of larger farms are more likely to consider themselves as farmers and respond to the census. Another reason for more complete coverage of the large farm group is because of the more intensive follow-up of farms with sales of \$100,000 or more. Census coverage in the Midwest region was substantially more complete than in the Northeast, South, and West due primarily to the higher proportion of larger farms in the Midwest region.

3.3 Characteristics of Undercounted Farms

The farms not on the mail list (10.6 percent), plus the misclassified farms (3.1 percent), are equal to the undercounted farms (13.7 percent). Three out of four of the undercounted farms were not on the mail list and one out of four was misclassified.

The preliminary coverage estimates indicate that 75.6 percent of the undercounted farms were

small operations with less than \$2,500 in sales of agricultural products. Another indication of their relatively small size is that while the undercounted farm rate is 13.7 percent, the undercounted farms accounted for only 1.9 percent of the estimated value of agricultural products sold in the United States in 1982. Acres on undercounted farms accounted for 1.7 percent of the estimated acres in the United States. The average size of undercounted farms was 66 acres compared to 439 acres for farms in the census.

An intensive analysis will be completed to determine additional characteristics of the undercounted farms, such as major crops and live-stock, land in farms, value of sales, and operator characteristics for the final coverage

evaluation publication. Additional research on alternative data collection and processing procedures will be conducted in order to decrease the number of undercounted farms resulting from misclassification and mail list deficiencies in future censuses.

3.4 Accuracy of the Estimates

Estimates of sampling variability expressed as estimated relative standard errors are presented in Table 3. The estimated relative standard error for a statistic is derived by dividing the estimated standard error for a statistic by that statistic. The estimated relative standard error for the proportion of farms not on the mail list at the U.S. level is 13.2 percent and ranges from

TABLE 3. Preliminary Estimates of Census Coverage for the 1982 Census of Agriculture by Value of Sales

	FARMS NOT ON MAIL LIST		MISCLASSIFIED FARMS		OVERCOUNTED FARMS	
	Estimated Percent	Relative Standard Error (Percent)	Estimated Percent	Relative Standard Error (Percent)	Estimated Percent	Relative Standard Error (Percent)
	(1)	(2)	(3)	(4)	(5)	(6)
<u>UNITED STATES</u>						
Total Farms	10.6	13.2	3.1	19.0	4.6	27.1
Farms by Value of Sales						
Less than \$2,500	28.5	12.8	6.8	25.8	6.8	41.1
\$2,500 or more	2.7	26.0	1.5	20.1	3.7	37.0
<u>NORTHEAST</u>						
Total Farms	20.3	18.6	2.0	24.0	3.7	35.2
Farms by Value of Sales						
Less than \$2,500	40.5	18.4	3.7	30.1	3.7	88.6
\$2,500 or more	8.5	29.8	0.9	44.4	3.7	55.3
<u>MIDWEST ^{1/}</u>						
Total Farms	4.2	44.1	2.6	21.7	3.6	53.5
Farms by Value of Sales						
Less than \$2,500	16.0	49.3	8.1	28.3	3.0	52.7
\$2,500 or more	1.7	72.8	1.4	32.9	3.8	61.8
<u>SOUTH</u>						
Total Farms	13.7	19.3	3.7	35.3	6.2	38.2
Farms by Value of Sales						
Less than \$2,500	30.9	18.2	7.2	43.4	9.0	55.6
\$2,500 or more	2.3	38.9	1.4	37.3	4.3	53.2
<u>WEST ^{2/}</u>						
Total Farms	15.0	19.7	3.4	18.0	3.1	56.2
Farms by Value of Sales						
Less than \$2,500	32.6	20.0	5.4	24.5	6.0	80.4
\$2,500 or more	4.8	36.9	2.2	28.2	1.5	43.9

^{1/} The Midwest Region was designated as the North Central Region until June 1984.

^{2/} Excludes Alaska and Hawaii.

Note: Net Coverage Rate (Percent) = Column (1) + Column (3) - Column (5).

18.6 percent in the Northeast to 44.1 percent in the Midwest. The estimated relative standard error for the net undercounted farm rate (9.1 percent) is 22.7 percent at the U.S. level. Estimated standard errors for regional estimates are high for some estimates and the data should be used with caution.

Two types of errors are possible in estimates based on a sample--sampling error and nonsampling error. Sampling error occurs because observations are made only on a sample and not the entire population. The estimated relative standard errors in Table 3 represent only the variation due to sampling error. Nonsampling error includes all remaining error and can be attributed to many sources: inability to obtain information about all cases in the sample (nonrespondents, refusals, incomplete report forms), definitional difficulties, differences in the interpretation of questions, incorrect information provided by the respondents, mistakes in recording or coding the data obtained, and other errors of collection, response, processing, coverage, and estimation for missing data. The "accuracy" of a survey result is determined by the joint effects of sampling and nonsampling errors.

4. SUMMARY

The results of the coverage evaluation indicate the continuing problem of census undercoverage. The undercounted farms in the census are primarily small units not likely to be included in any of the administrative source lists used to develop the census mail list. The major statistics affected by undercoverage in the census are the number of farms and averages based on the number of farms. Undercoverage has little effect on the total value of agricultural products sold, acres, and other related data. The undercoverage of small farms has varying significance to census user groups depending on their data needs. The

group of misclassified farms offers an area for potential improvement through additional research in mail list development and census processing.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the assistance of Margaret Bruce, Eileen Gray, and Kathleen Campbell for their invaluable manuscript word processing, and a number of colleagues for their constructive comments.

REFERENCES

1. Dea, Jane Y., Tommy W. Gaulden, and D. Dean Prochaska, "Record Linkage For The 1982 Census Of Agriculture Mail List Development Using Multiple Sources," 1984 Proceedings of the Section on Survey Research Methods, American Statistical Association.
2. Ruggles, Donna R., Jane Y. Dea, Flora Kwok, and Cindy A. Carman, "Evaluation Of The Effectiveness Of Data Collection Procedures For The 1982 Census Of Agriculture," 1984 Proceedings of the Section on Survey Research Methods, American Statistical Association.
3. 1978 Census of Agriculture, Volume 5--Special Reports, Part 3--Coverage Evaluation.
4. Chapman, David D., and Charles E. Rogers, "1978 Census Of Agriculture--Area Sample Design And Methodology," 1978 Proceedings of the Survey Research Section, American Statistical Association.
5. Wolter, Kirk M., "Coverage Error Models For Census And Survey Data," 1983 Proceedings of the International Statistical Institute, Madrid, Spain.