

Volume 5

Special Reports

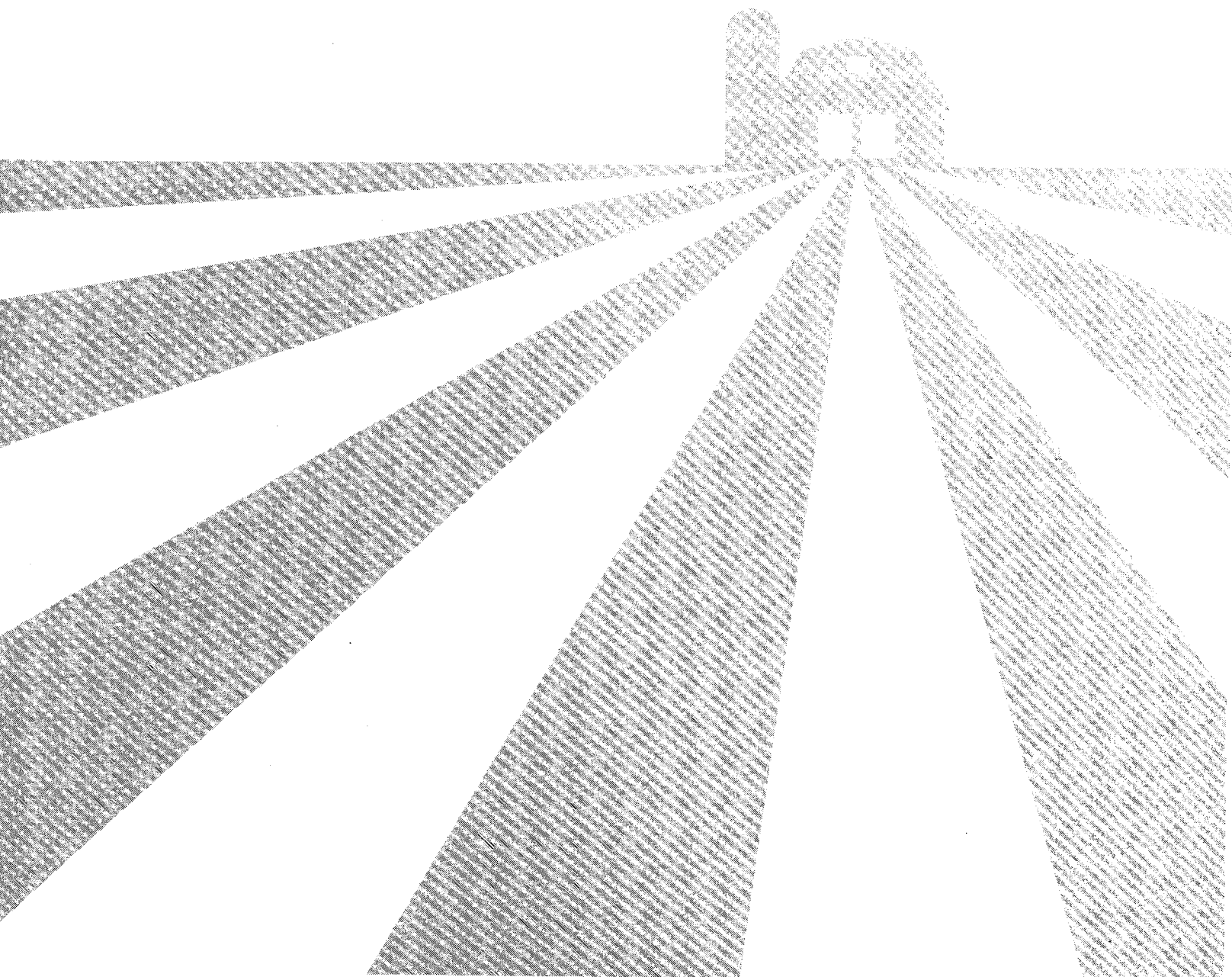
Part 3

**Coverage
Evaluation**

AC78-SR-3

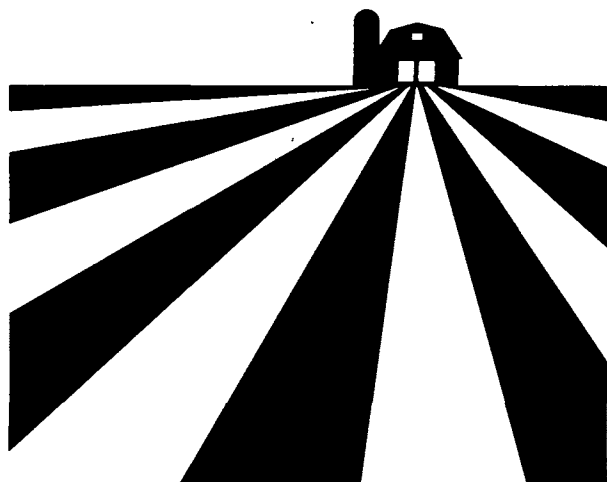
1978
CENSUS OF
AGRICULTURE

U.S. Department of Commerce
BUREAU OF THE CENSUS



1978 CENSUS OF AGRICULTURE

Volume 5
Special Reports



Part 3
**Coverage
Evaluation**

AC78-SR-3

Issued May 1982



U.S. Department of Commerce
Malcolm Baldrige, Secretary
Robert G. Dederick,
Assistant Secretary for Economic Affairs

BUREAU OF THE CENSUS
Bruce Chapman,
Director



BUREAU OF THE CENSUS

Bruce Chapman, Director

**Shirley Kallek, Associate Director for
Economic Fields**

**Michael G. Farrell, Assistant Director for
Economic and Agriculture Censuses**

AGRICULTURE DIVISION

George Pierce, Acting Chief

ACKNOWLEDGMENTS—Many persons contributed to the coverage evaluation program for the 1978 Census of Agriculture. First and foremost are the many farm and ranch operators, members of their families, and employees who furnished the information.

Primary direction was by **Shirley Kallek**, Associate Director for Economic Fields, and **Richard B. Quanrud**, Assistant Director for Economic and Agriculture Censuses (to June 1981), and **Michael G. Farrell**, his successor. Technical direction and guidance were provided by **Barbara A. Bailer**, Associate Director for Statistical Standards and Methodology.

The coverage evaluation program was prepared in the Agriculture Division under the direction of **Orvin L. Wilhite**, Chief (to January 1980), **Arnold L. Bollenbacher**, his successor (to April 1982), and **Charles E. Rogers**, Assistant Chief for Research and Methodology (to June 1981).

The planning, processing, analyses, and tabulation of the statistics were performed under the supervision of **D. Dean Prochaska**, Chief for the Program Research and Development Branch. **Jane Y. Dea** was responsible for the overall planning and coordination of the various phases. Persons providing major assistance in the processing, analyses, tabulation, and publication phases were **Dorothy M. Blackstone**, **William C. Davie**, and **Emily Lorenzen**. Persons providing technical and clerical assistance were **James A. Aldeman**, **Melody M. Atkinson**, **Margaret I. Bruce**, **James Joerger**, **Diane Kenney**, **Mary F. Marbury**, **Diana Marz**, **Donna M. Rafferty**, **Marjorie R. Turner**, and **Antoinette P. Wooten**.

Sampling error data were provided by **David D. Chapman**, Agriculture Division, and **George F. Train**, Statistical Methods Division.

Computer programming was performed by **Jeffrey L. Brubaker** and **William M. Sandusky**, under the direction of **Lowell T. Wrucke**.

Cecelia H. Peets, **Regina E. Lindsey**, and **Karin G. Beach** were responsible for the various phases of the publication process.

Important contributions were made by many individuals from the following divisions: Administrative Services—forms design; Data Preparation—data keying, survey materials distribution, and map preparation; Demographic Surveys—coordination and development for the Annual Housing Survey Agriculture Supplement; Field—data collection and followup; and Publications Services—publication planning and design, editorial review, composition, and printing procurement.

Library of Congress Cataloging in Publication Data

United States. Bureau of the Census.
1978 census of agriculture.

Supt. of Docs. no.: L3.31/

CONTENTS: v. 1. State and county data.—v. 2. Statistics by subject.—v. 3. Agriculture services. [etc.]

1. Agriculture—United States—Statistics. I. Title. II. Title: Census of agriculture.

HD1753 1978.B87 7 1980 338.1'0973

79-600215

For sale by Data User Services Division, Customer Services (Publications), Bureau of the Census, Washington, D.C. 20233, or any U.S. Department of Commerce district office. Postage stamps not acceptable; currency submitted at sender's risk. Remittances from foreign countries must be by international money order or by a draft on a U.S. bank. \$2.25 per copy.

CONTENTS

	Page
Introduction.	IV
General Explanation	V
MAP—Regions of the United States.	XIII

TABLES

1. Farms by Sales Group, Standard Industrial Classification, Size, and Operator Characteristics, by Components of Coverage: 1978	1
2. Characteristics of Missed Farms by Sales Group: 1978	3
3. Land in Farms by Sales Group and Components of Coverage: 1978	4
4. Value of Agricultural Products Sold by Sales Group and Components of Coverage: 1978	4
5. Reliability of Farm Estimates by Sales Group, Size of Farm, and Components of Coverage: 1978.	5

APPENDIXES

A. Farms Adjusted for Undercount: 1978 and 1974.	A-1
B. Report Forms.	B-1

INTRODUCTION

PURPOSE OF CENSUS COVERAGE EVALUATION	Page IV
HISTORY AND AUTHORITY	IV
SYMBOLS	IV

HISTORY AND AUTHORITY

The 1978 Census of Agriculture was the 21st nationwide agriculture census conducted in the United States. The first agriculture census was taken in 1840 as part of the sixth decennial census of population. From 1840 to 1920 an agriculture census was taken every 10 years. Beginning in 1925, the census of agriculture was conducted every 5 years. In 1976, Congress authorized the census of agriculture to be taken for 1978 and 1982 and every 5 years thereafter to coincide with the economic censuses. The census of agriculture is taken in accordance with the provisions of title 13, United States Code.

SYMBOLS

The following symbol is used throughout the tables.

— Zero.

PURPOSE OF CENSUS COVERAGE EVALUATION

The Bureau of the Census seeks to measure the accuracy and completeness of its statistics for each census of agriculture through a coverage evaluation program. This program provides an independent check on the results and provides information to identify problem areas associated with coverage errors as a basis for developing improvements for future censuses. The results from this program serve as an important means of informing the users of the data of any known deficiencies which might affect their interpretation and uses of the data.

GENERAL EXPLANATION

	Page
1978 CENSUS OF AGRICULTURE	V
1978 CENSUS COVERAGE EVALUATION	VI

1978 CENSUS OF AGRICULTURE

Farm Definition

It has been necessary to establish minimum criteria for defining a farm for census purposes. The farm definition has been changed nine times since 1840, as the Nation's agricultural economy has grown and changed. The current farm definition, first used for the 1974 census, is any place from which \$1,000 or more of agricultural products were sold or potentially could have been sold during the census year. The previous definition was any place with less than 10 acres from which \$250 or more of agricultural products were sold or potentially could have been sold during the census year, or any place of 10 acres or more from which \$50 or more of agricultural products were sold or potentially could have been sold during the census year. In all censuses, places not having sufficient sales to qualify as a farm could qualify on potential sales, based on the production of crops and/or livestock which were not sold.

Data Collection

Before 1969, the census of agriculture was based on a canvass of rural areas by enumerators and personal interviews of farm operators. The 1969 and 1974 censuses used a mailout, self-enumeration, mailback procedure to collect the data. In 1978, the mailout/mailback procedure was supplemented by the Census of Agriculture Area Sample (CAAS), a direct enumeration sample.

Mail List

The 1978 mail list was assembled from the 1974 census farm list and from records obtained from the Internal Revenue Service, the U.S. Department of Agriculture, other government agencies, and agriculture-related organizations. Lists of large and specialized operations were obtained from various trade associations and other sources. The total number of records from all sources was about 17.5 million.

Because a name and address could appear on more than one source list, a record linkage process was used to remove duplicates from the preliminary list. Records on the preliminary list that were not likely to be farms (based on the source list or lists they appeared on) were included in the 1978 Farm and Ranch Identification Survey. These addressees were mailed a short screening questionnaire to identify their current status. As a result of the Farm and Ranch Survey, addressees with no agricultural operations were excluded and new tenants and suc-

cessors were added. The final census mail list contained approximately 4.4 million names and addresses.

Census of Agriculture Area Sample (CAAS)

Because previous coverage evaluation studies had shown that many smaller farms were not included on the census mail list, CAAS was used to improve the completeness of farm coverage. CAAS contained approximately 6,400 segments in rural areas (areas with less than 2,500 population). Enumerators canvassed all households in the segments in October and November 1978 and completed a census report form for each agricultural operation. These forms were matched to the census mail list. The forms for addresses that were not matched were processed as CAAS additions. Data from the final nonmatched report forms were used to estimate the number and characteristics of farms not on the mail list at the State, regional, and national levels. No county level estimates from CAAS were developed because the sample size was insufficient to provide reliable estimates at that level.

Report Forms

Two report forms were used in the mail portion of the census. A five-page form containing all the census items was sent to all large farms (based on expected sales and/or acres), farms with special characteristics, and farms from samples of other addresses. A four-page form omitting sample items was sent to the remaining addresses. The form used in the CAAS was similar to the five-page form but included additional items used for matching names and addresses to the mail list.

Mailing and Followup

Report forms were mailed in late December 1978. Nonrespondents were sent a series of seven followup letters—three with report forms—at 3- to 4-week intervals. Additional mailings and telephone calls were made in low response areas. Telephone calls were made to all nonrespondents who were expected to have large operations. The final response rate for the 1978 Census of Agriculture was about 88 percent. A nonresponse adjustment procedure was used to represent the final nonrespondent farms in the census results. This component of the census farm count is subject to sampling variability. A description of the nonresponse adjustment procedure is included in volume 1, appendix A.

Processing Procedure

Completed report forms were clerically reviewed prior to data keying. After keying, a detailed item-by-item computer consistency edit of the data was performed. Problem forms were subjected to a special review. In some cases, telephone calls were

made to respondents to resolve conflicting data or provide missing information.

After each form was screened, computer edited, and all identified problems corrected, the data items were tabulated by computer. The tabulated totals were reviewed county by county to identify and correct any remaining problems before publication.

1978 CENSUS COVERAGE EVALUATION

History

Coverage evaluation studies have been conducted for each census of agriculture since 1945. The basic procedures for each study have been:

- a. Select an area probability sample, or use an existing sample and identify the farms in the sample to establish a measurement base which represents the universe of farms.
- b. Match the sample farms to the census farms to determine the relationship of the census to the base sample units.
- c. Followup to check and clarify differences.
- d. Process, tabulate, analyze, and publish the results.

The results of previous coverage evaluation studies have influenced census procedures. Before 1950, interviewers were given the farm definition and instructed to enumerate all places conforming to that definition. The 1945 study showed that marginal farming operations were a large proportion of the missed farms. To improve the coverage of these marginal operations, the enumeration procedures were changed. Starting in 1950, interviewers were instructed to enumerate all places with specified agricultural activities and the farm definition was applied during processing.

In 1954, two new techniques were introduced to reduce the undercoverage of farms. A township sketch was used in selected counties to improve the coverage of nonresident operators, that were shown by the 1950 evaluation study to be a large share of the missed farms. Enumerators in these counties were required to draw the boundaries of each farm and each nonfarm tract on the township sketch. In addition, a listing book was used to record the location and identification of every dwelling and every place with no dwelling, but with agriculture operations, in each enumeration district.

The 1964 coverage evaluation study found that undercoverage of small farms was a continuing problem. Other studies indicated that at least equal and perhaps better coverage could be obtained with a mailout/mailback procedure. The mailout/mailback method of data collection was first used in the 1969 census and has been used in all subsequent censuses.

Coverage evaluation for 1969 and 1974 indicated that the source lists acquired for data collection by mail did not provide adequate coverage of smaller farms. In 1969 and 1974, 33 percent and 27 percent respectively, of all census farms less than \$2,500 were missed. Because of this, the 1978 census was supplemented by CAAS, a coverage improvement survey designed to increase coverage of small farms at the State level. The area sample provides State, regional, and national levels.

Objectives

The 1978 coverage evaluation program was planned and developed with a specific set of objectives. These objectives were developed from a review of previous census coverage evaluation objectives along with the addition of several new objectives aimed at determining the effectiveness of new procedures developed for the 1978 Census of Agriculture. The specific planned objectives were:

- a. Provide measures of the accuracy of census farm counts and a limited number of other items, such as land in farms and value of agricultural products sold, to aid users in the interpretation and utilization of the data.
- b. Provide estimates indicating the characteristics of missed farms.
- c. Provide information relating to the accuracy of the census area sample estimates and potential problem areas associated with the enumeration procedures.
- d. Provide information on factors associated with census error and identify problem areas to improve coverage in future censuses.

Sample Survey Design and Methodology

The coverage evaluation program for the 1978 Census of Agriculture was based on two surveys—the Annual Housing Survey (AHS) and the Post Enumeration Survey (PES).

Annual Housing Survey

The AHS is an ongoing survey conducted by the Bureau of the Census used primarily to measure household characteristics. The sample represents all households in the United States.

The 1978 sample consisted of delineated sample areas spread geographically across the 50 States. About 72,000 housing units, both occupied and vacant, were included. The overall sampling rate in urban areas with population 2,500 or more was 1 in 1,366. For rural areas, with less than 2,500 population, the sampling rate was 1 in 683.

The AHS agriculture supplement (see appendix B) consisted of a short series of agriculture screening questions added to the report form of the 1978 AHS to identify potential census farms. Alaska and Hawaii were excluded from the coverage evaluation sample.

The farm universe identified in AHS was used primarily to estimate the number and the characteristics of farms not on the mail list, and operated by individuals living in urban areas with population 2,500 or more. These areas were not covered by CAAS. Also, the AHS provided measures of error for misclassified farms on the mail list. The AHS was completely independent from the 1978 Census of Agriculture, from the standpoint of data collection and source lists.

The principal processing steps for the AHS agricultural supplement were:

- a. Receive AHS supplements from the census regional offices following field enumeration.
- b. Identify supplements with potential agricultural activities.

- c. Match supplements with potential agricultural activities to the census microfilm mail file. Identify match or doubtful match cases.
- d. Mail evaluation report form (78-A90; see appendix B) to nonmatch and doubtful match cases to obtain basic farm data and additional information for matching.
- e. Mail three followups to nonrespondents at 2-week intervals.
- f. Telephone followup of remaining nonrespondents by regional offices.
- g. Perform second match to mail list using information from form 78-A90.
- h. Review form 78-A90 to classify as farm/nonfarm according to farm definition and match/nonmatch to census mail list.
- i. Telephone followup to resolve problem cases and obtain form 78-A90 for noncontact cases.
- j. Assign AHS weights and population size codes to all cases.
- k. Obtain census data records from data file for all match cases for use in assigning farm and coverage classification codes.
- l. Review and assign coverage classification codes to all cases.
- m. Key data to produce coverage data file.
- n. Edit and review data for accuracy and consistency.
- o. Produce estimates for AHS sample error.
- p. Tabulate data and publish results.

Post Enumeration Survey

The PES was a personal interview survey conducted in December 1978 following the completion of the CAAS enumeration. Its purpose was to collect data for evaluating the area sample portion of the 1978 Census of Agriculture. The PES sample consisted of a 1 in 30 subsample of the original 6,391 area sample segments selected systematically across the conterminous States. The 212 segments enumerated had an average of 75 households per segment.

The PES was conducted by field enumerators under the supervision of the Census Field Division regional offices. Only highly qualified enumerators were selected based on past performance evaluations. Intensified canvassing methods, probing interview techniques, and special forms were utilized to achieve the best coverage and quality possible. Interviews were conducted with the most knowledgeable person in each household. All members of the household were listed in the PES listing book (see appendix B), whereas, only the head of the household was listed in CAAS. A series of screening questions was used to determine which household members had agricultural operations. As a last resort, respondents who could not be personally interviewed were interviewed by telephone.

The principal processing steps were:

- a. Receive PES listing books from census regional offices following PES field enumeration. Procure CAAS materials for PES subsampled segments.

- b. Match PES households to CAAS households and identify PES cases with potential agricultural activities that are nonmatch to CAAS.
- c. Match PES cases with potential agricultural activities that are nonmatch to CAAS to the census microfilm mail file.
- d. Mail evaluation report form (78-A90; see appendix B) to nonmatch and doubtful match cases to obtain basic farm data and additional information for matching.
- e. Mail three followups to nonrespondents at 2-week intervals.
- f. Telephone followup of remaining nonrespondents by regional offices.
- g. Perform second match to mail list using information from form 78-A90.
- h. Review form 78-A90 to classify as farm/nonfarm according to farm definition and match/nonmatch to CAAS and census mail list.
- i. Telephone followup to resolve problem cases and obtain a form 78-A90 for noncontact cases.
- j. Assign area sample segment weights to all cases.
- k. Obtain census data records from data file for all match PES cases for use in assigning coverage classification codes.
- l. Review and assign coverage classification codes to all cases.
- m. Key data to produce coverage data file.
- n. Edit and review data for accuracy and consistency.
- o. Produce estimates for PES sample error.
- p. Tabulate data and publish results.

Results from the PES were used to provide a statistical basis for evaluating the completeness of the area sample results.

Estimation Procedure

The coverage evaluation provides estimates of three coverage components in relation to the census:

- a. Included farms.
- b. Overcounted farms.
- c. Missed farms.

The estimates are based on the AHS and PES samples and take the general form, $Y_t = Y_i - Y_o + Y_m$ where:

Y_t = Estimate of total farms as determined in the coverage evaluation.

Y_i = Estimate of all farms included in the census.

Y_o = Estimate of farms overcounted in the census.

Y_m = Estimate of farms missed in the census.

The estimates of proportion of farms included in the census are in the form,

$$\text{included (percent)} = (Y_i/Y_t) \times 100.$$

The estimates of the proportion of net missed farms are in the form,

$$\text{net missed (percent)} = \frac{(Y_m - Y_o)}{Y_t} \times 100.$$

Results

Estimates of Census Coverage

Estimates of census coverage of farms were made only at regional and national levels since evaluation samples were too small to provide reliable estimates at State or county levels. Estimates of the number of farms are based on a combination of the AHS and PES samples. Estimates for the value of agricultural products sold and land in farms are based on the AHS and PES sample estimates for the included, overcounted, and missed farms. The estimates for land in farms and value of agricultural products sold do not represent a measurement of total error for these items since reporting error was not measured for included and overcounted farms.

The estimates produced in the coverage evaluation program should be considered relative to the census economic data as well as the farm count. Estimates of the total number of missed farms or the proportion of missed farms alone, are not a complete indication of the quality of the census. Consideration of economic characteristics such as estimates of the value of agricultural products sold along with the farm counts may be a better indication of census quality and, in turn, may have a greater impact on the user's needs. For example, while the net missed farm rate was 3.4 percent, the missed farms accounted for only 1.6 percent of the estimated value of agricultural products sold in the United States.

Regional estimates are presented in Tables 1, 3, and 4 to provide some indication of census coverage below the national level. Because of the relatively high sampling error these estimates may not be reliable. Caution should be observed when drawing conclusions based upon comparisons of regional estimates within and between tables.

Table 1—This table presents the number of farms by sales group, standard industrial classification (SIC), size, and operator characteristics by components of coverage. Farms were classified as included, overcounted, and missed. Overcounted farms were part of the farms included in the census. Estimates indicate that 96.6 percent of all farms were included in the 1978 census for the conterminous United States. Approximately 4.4 percent of all farms were missed and approximately 1.1 percent of all farms were overcounted resulting in an average net missed rate of approximately 3.4 percent for data at the State level and above. The average net missed rate was 15.0 percent in 1969 and 10.7 percent in 1974 for data at all levels. Comparison of these rates indicates the considerable improvement for data at the State level and above provided primarily by the inclusion of the area sample with the 1978 net missed rate being reduced to 3.4 percent from 10.7 percent in 1974. In 1978, the area sample estimates were not included in census county data.

For farms with value of agricultural products sold of \$2,500 or more, 97.6 percent were included in the census. The net missed farm rate for this group was 2.4 percent. Larger farms were more likely to be included in census source lists, and received more intensive followup and processing to ensure that they were included.

An estimated 93.5 percent of farms with value of agricultural products sold of less than \$2,500 were included in the census. The net missed farm rate was 6.5 percent. The net missed rate for this group was 31.6 percent in 1969 and 25.9 percent in 1974. Coverage of small farms was improved primarily by use of CAAS and by changes in the development of the mail list.

The estimated number of overcounted farms was approximately 24,000. Overcounting occurred primarily when census reports were duplicated for a single farm or when multiple census reports were included for parts of a single farm. In addition, overcounting occurred when a nonfarm was counted as a census farm or when a farm was incorrectly classified in the area sample.

Farms missed in the census are classified into three groups:

- Group 1. Farms on the mail list which were misclassified as nonfarms because of incorrect reporting, incomplete reporting, and processing errors.
- Group 2. Farms in urban areas excluded from the area sample and not located on the census mail list.
- Group 3. Farms missed in CAAS and not located on the census mail list.

About 57 percent of the missed farms were misclassified, about 21 percent were missed in urban areas, and about 22 percent were missed in CAAS.

While about 62 percent of the missed farms had value of agricultural products sold of \$2,500 or more, only 10 percent were "large" farms with sales of \$40,000 or more. About 64 percent of the missed farms had less than 100 acres, and only 7 percent had 500 acres or more. Of the missed farms, 66.6 percent were operated by full owners, 12.2 percent by part owners, and 21.2 percent by tenants. Missed farms were divided equally between livestock farms and crop farms.

The net missed rate for nonresident operators in 1978 was 7.7 percent. Coverage of operators not living on their farms has been a problem in past censuses because of the difficulty in enumerating operators living in urban areas or in small towns away from their farms. Various procedures have been introduced in previous censuses to attempt to improve enumeration of nonresident operators. Although some improvement in the coverage of nonresidents has been made over the years, a relatively high undercoverage rate remains.

Table 2—This table presents the characteristics of missed farms by sales group. The missed farm data do not represent total error in the census because detailed data for the overcounted farms could not be derived in the coverage evaluation and reporting error on correctly counted farms was not measured.

The estimated total number of missed farms was approximately 101,000 or 4.4 percent of the estimated total number of farms. The average size of missed farms was 202 acres.

Table A presents selected characteristics of missed farms compared to census totals. Sample estimates of missed farm characteristics were not developed for the coverage samples and comparisons for these items can be made only by using census totals. While these estimates probably understate the total error, the missed farm estimates for these items are likely

to contribute substantially more than other components to total error. Therefore, estimated minimum levels are indicated by adding the missed farm estimates to corresponding census figures for comparison. The data in table A provide some indication of census coverage for specified items.

Table 3—This table presents estimates of the land in farms by sales group and components of coverage. For the United States, an estimated 98.0 percent of the land in farms was included in the census. Missed farms accounted for only 2.0 percent of the land in farms. Missed farms accounted for 1.9 percent of the estimated total acres for farms with sales of \$2,500 or more, and 5.1 percent for farms with sales of less than \$2,500. The estimates for land in farms do not represent total error because reporting error was not measured on included and overcounted farms. No sampling errors were calculated for land in farms. However, estimates should be used with caution because relatively high sampling errors are likely.

Table 4—This table presents the estimates for the value of agricultural products sold by sales group and components of coverage. Estimates indicate that 98.4 percent of the value of agricultural products sold was included in the 1978 census for the conterminous United States. Missed farms accounted for 1.6 percent of the estimated value of agricultural products sold for farms with sales of \$2,500 or more, and 6.5 percent for farms with sales of less than \$2,500. The estimates for value of agricultural products sold do not represent total error because reporting error was not measured on included and overcounted farms. No sampling errors were calculated for value of agriculture products sold. However, estimates should be used with caution because relatively high sampling errors are likely.

Table 5—This table presents the reliability of farm estimates by sales group, size of farm, and components of coverage. Standard

errors were computed directly for the estimated total farms, included farms, and missed farms. The estimates of sampling error for the overcounted farms would have been based on a small number of observations and were not produced. Standard errors for regional estimates by value of sales and size of farm are high for some estimates and should be used with caution.

The relative standard error for the estimated total farms in the United States was 4.5 percent. The standard error for the estimate of included farms, as percent of estimated total, was 4.6 percent at the U.S. level, and ranges from 7.3 to 11.6 percent at the regional level. The relative standard error for missed farms was 11.2 percent at the U.S. level. Additional detail regarding sampling error may be found in the Accuracy of the Estimates section.

Other Results

One of the objectives of the 1978 coverage evaluation was to attempt to identify potential problem areas associated with the CAAS. The CAAS was used to supplement the mail list and provided approximately 8.9 percent of the total census farms at the U.S. level. Since CAAS was designed to cover rural areas only (areas with less than 2,500 population) the coverage evaluation studies were developed to provide estimates of farms operated by individuals living in urban areas (places with 2,500 inhabitants or more), as well as farms in rural areas. The AHS sample represented all population size areas; therefore, it provided the capability for the measurement of farms operated by individuals living in urban and rural areas. Estimates of missed farms in the census in urban areas are shown in table 1.

Table B shows estimates of the total number of farms by population of the area in which the residence of the operator is located. These estimates reflect the location of the farm operator household and not necessarily the location of the actual farm operation. The population areas are based on the 1970 population census information. The estimates for farms by population of area from the coverage evaluation samples are not comparable with census published data because of sampling and nonsampling errors. See, Accuracy of the Estimates.

To provide information on factors associated with census error and to identify problem areas to improve coverage in the future, three investigative studies were undertaken. These studies used information from the AHS and PES samples. The areas of study were:

- Farms on the mail list which were misclassified as non-farms in the census.
- Farms missed in CAAS.
- Farms overcounted in CAAS.

Table A. Census Farms and Estimates of Missed Farm Characteristics for Selected Items

(Data for Alaska and Hawaii are not included)

	Census farms published	Estimate for missed farms	Combined census farms published and missed farms	Ratio of missed farms to com- bined
Corn for grain.....farms..	842,894	20,232	863,126	2.3
acres..	70,733,245	1,042,520	71,775,765	1.5
Sorghum for grain..farms..	115,139	1,472	116,611	1.3
acres..	12,961,799	139,840	13,101,639	1.1
Wheat.....farms..	383,357	8,832	392,189	2.3
acres..	54,457,748	473,984	54,931,732	.9
Soybeans.....farms..	550,640	12,436	562,976	2.2
acres..	61,832,897	533,196	62,366,093	.9
Hay.....farms..	1,200,314	19,236	1,219,550	1.6
acres..	61,740,582	816,222	62,556,804	1.3
Tobacco.....farms..	203,015	5,888	208,903	2.8
acres..	1,004,697	39,229	1,043,926	3.8
Inventory:				
Cattle and calves.farms..	1,460,964	49,320	1,510,284	3.3
number..	105,487,755	2,060,076	107,547,831	1.9
Hogs and pigs....farms..	511,838	11,922	523,760	2.3
number..	58,759,075	193,340	58,952,415	.3
Hens and pullets.farms..	315,057	13,228	328,285	4.0
number..	357,787,310	264,054	358,051,364	.1

Table B. Farms by Population of Area

	Estimated farms	Percent
United States.....	2,279,470	100.0
1970 population of area:		
Less than 2,500 inhabitants.....	2,107,445	92.4
2,500 to 9,999 inhabitants.....	91,065	4.0
10,000 to 24,999 inhabitants.....	23,552	1.0
25,000 to 99,999 inhabitants.....	35,328	1.6
100,000 inhabitants or more.....	22,080	1.0

Exact causes of census error could not always be determined. However, in each study, all available information was used to reach reasonable conclusions.

A special study was initiated to determine reasons for misclassification of farms. Misclassification accounted for 57 percent of the missed farms because of incorrect or incomplete reporting or processing errors. The available information on each misclassified case (form 78-A90, telephone followup materials, materials from the match to the mail list, etc.) was carefully reviewed and additional matching to the mail list was conducted if any information had been overlooked in the initial matching process. A final review determined the possible reason why each case had been misclassified.

The results of the study showed that there were varied reasons for misclassification in the census. It appears that the major reason was that some census respondents felt that their operations were "too small" or "only for home use" and should not be classified as a farm. Therefore, these respondents did not report any agricultural activities or failed to report the full extent of their activities. Some changes in the design of the report form and additional review of nonfarm report forms were suggested to reduce the problem of misclassification in future censuses.

A second special study attempted to determine why farms had been missed in CAAS. The PES missed farm cases, the CAAS and PES listing books, maps, and all other information were carefully reviewed.

The conclusions reached by the study were: (1) differences in reporting data often arose when the respondent was someone other than the farm operator; (2) enumerators need to check more thoroughly for households in isolated locations so that all households in the segment are covered; and (3) smaller farm operations are most often missed because the operators feel their operations are too small to be classified as a farm and sometimes fail to give the enumerator, even after probing, complete information. It was recommended that the CAAS enumerator's instructions be modified so that the enumerators rely less on neighbors or other persons outside the household for agricultural information on the household.

The objective of the third special study was to try to determine why farms were overcounted in CAAS. Overcount occurred when a CAAS farm should have been matched to the census mail list but was incorrectly classified as a nonmatch; therefore, data from the same farm was included in both CAAS and the census. A thorough review of the CAAS and PES listing books and all information available on the PES overcount cases was completed. Farm data from CAAS and the census were compared to determine if there was duplication.

Analysis indicated that the three recurrent problems during the CAAS matching operation which led to overcount in CAAS were: (1) misspelled names in CAAS and/or the census mail list; (2) alternate addresses for the same operation; and (3) alternate names for the same operation. Additional review of the report forms for alternate names and addresses and changes in the matching procedures were recommended. Also, it was recommended that the name of the spouse be included on the CAAS report form for use in the matching procedure.

Accuracy of the Estimates

The statistics in this report are estimates derived from AHS and PES coverage evaluation sample data. 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. Nonsampling error includes all remaining error and can be attributed to many sources, such as inability to obtain data for all cases in the sample, response error, definitional differences, coding errors, processing problems, interviewer interpretation, and analyst effects. The "accuracy" of a survey result is determined by the joint effects of sampling and nonsampling errors.

Sampling error—The sample used in this survey was one of a large number of possible samples of the same size that could have been selected using the same sample design. Estimates derived from the different samples would differ. The deviation of a sample estimate from the average of all possible samples is called the sampling error. The standard error of a survey estimate is a measure of the variation among the estimates from the possible samples and thus is a measure of the precision with which an estimate from a particular sample approximates the average result of all possible samples. The relative standard error is defined as the standard error divided by the value being estimated.

The sample estimates and the estimates of absolute and relative standard errors presented in table 5 permit the construction of interval estimates with prescribed confidence that the interval includes the average result of all possible samples.

If all possible samples were selected, each of these surveyed under essentially the same conditions and an estimate and its estimated relative standard error were calculated from each sample, then:

- a. Approximately 67 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average value of all possible samples.
- b. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average value of all possible samples.

For example, the estimated total number of farms in the United States is 2,279,470 with a relative standard error of 4.5 percent. The standard error is 102,576 (4.5 percent of 2,279,470) and the chances are 2 out of 3 (67 percent) that complete coverage using the same survey methods would yield between 2,176,894 and 2,382,046 farms.

As calculated, the standard error also partially measures the effect of nonsampling errors but does not measure the effect of any systematic biases in the data arising from incorrect reporting by respondents, adjustments for nonresponse, duplication, or incomplete coverage of farms.

The following examples describe how the published standard errors in table 5 may be used to calculate approximate standard errors for selected types of derived statistics, such as relative

standard errors of ratios (example 1) and standard errors of differences between ratios (example 2).

The relative standard errors, supplemented by the correlation coefficient, may be used to develop approximate standard errors for various estimated ratios, $R' = Y'/X'$, where Y' is a subset of X' .

Example 1—Approximate relative standard errors of ratios of different items for a given classification may be computed by the formula,

$$V(R') = \sqrt{V^2(Y') - 2\rho(Y', X')V(Y')V(X') + V^2(X')},$$

where $V(Y')$ and $V(X')$ are the relative standard errors of each of the two item totals, $V^2(Y')$ and $V^2(X')$ are the squares of those relative standard errors, and $\rho(Y', X')$ is the correlation coefficient between the estimates.

The correlation coefficient may be computed by the formula,

$$\rho(Y', X') = \frac{\sigma^2(Y')}{\sigma^2(X')},$$

where $\sigma^2(Y')$ is the square of the absolute standard error of Y' and $\sigma^2(X')$ is the square of the absolute standard error of X' .

To compute the standard error of the percent of missed farms in the North Central Region, which is the ratio of missed farms to total farms, from table 5, apply the formula shown above for relative standard errors of ratios.

$$\begin{aligned} Y' &= 37,132 \\ X' &= 938,248 \\ R' &= Y'/X' = .040 \\ V(Y') &= .215 \\ V(X') &= .076 \end{aligned}$$

$$\rho(Y', X') = .013$$

Substituting these values into the formula gives,

$$\begin{aligned} V(R') &\doteq \sqrt{.0462 - 2(.013)(.215)(.076) + .0058} \\ &\doteq \sqrt{.0516} \\ &\doteq .227 \end{aligned}$$

Therefore, the estimated ratio of 4.0 percent missed farms in the North Central Region is subject to a relative standard error of 22.7 percent.

Example 2—Approximate relative standard error for the difference between ratios,

$$D' = R_1' - R_2' \text{ where } R_1' = \frac{Y_1'}{X_1'} \text{ and } R_2' = \frac{Y_2'}{X_2'},$$

may be computed by the formula, $V(D') = V(R_1') + V(R_2')$, where $V(R_1')$ and $V(R_2')$ are the relative standard errors

of each of the ratios, assuming the two ratios to be statistically independent.

To compute the standard error of the difference of the ratio of missed farms in the North Central Region from the ratio of missed farms in the Northeast Region, apply the formula shown above for the relative standard error of the difference between ratios.

$$\begin{aligned} R_1' &= \text{ratio of missed farms in Northeast} = .064 \\ R_2' &= \text{ratio of missed farms in North Central} = .040 \\ D' &= .064 - .040 = .024 \\ V(R_1') &\doteq .322 \\ V(R_2') &\doteq .227 \\ \text{Substituting these values into the formula gives,} \\ V(D') &= .322 + .227 = .549 \end{aligned}$$

Therefore, the estimated difference of 2.4 percent is subject to a relative standard error of about 54.9 percent.

Nonsampling error—Approximately 72,000 sample housing units (both occupied and vacant) were eligible for interview in the 1978 AHS. Of this number, 6.1 percent or 4,400 units were classified as noninterviews. A unit was classified as noninterview if the occupants refused to be interviewed or could not be contacted after repeated visits. In addition, about 4.0 percent or 2,900 units were partially completed interviews with the agriculture supplements being classified as noninterview. The majority of the noninterview agriculture supplements were due to refusals. The total noninterview rate for the agriculture supplement was about 10 percent.

An additional factor contributing to possible nonsampling error in the coverage estimates is that about 5 percent of the total AHS supplements with agriculture were unclassified. Unclassified cases are those which could not be identified as either a farm or a nonfarm in the coverage evaluation processing. If the correct classification could have been determined, the unclassified group most likely would have been spread throughout all coverage components. However, it is likely that the unclassified group would be concentrated more heavily in the missed farms component since the majority of these cases were not matched to the mail list.

The assumption that all nonrespondent farms are correctly represented in the census as a result of the nonrespondent adjustment procedure may produce some bias in the coverage estimates. The nonrespondent adjustment represented about 8.5 percent of the farms and about 4 percent of the value of agricultural products sold in the 1978 census. The coverage sample had a 7.5 percent adjustment rate compared with the 8.5 percent adjustment rate in the census.

Variance estimation—Estimates in this evaluation study are the sum of two separate and statistically independent surveys—the AHS and the PES. Estimates of totals and their variances are the sum of the two separate survey estimates. The evaluation of totals and their sampling variances are discussed separately for each survey.

AHS—The 1978 AHS estimates are based on data collected in October 1978 through January 1979. The sample for this survey was spread over 461 sample areas (called primary sampling units), comprising 923 counties and independent cities with coverage in each of the 50 States and the District of Columbia.

To select the sample areas, the United States was divided into areas made up of counties and independent cities referred to as primary sampling units (PSU's). These PSU's were then grouped into 376 strata, 156 of which consisted of only 1 PSU in sample with certainty. These 156 strata, mostly the larger standard metropolitan statistical areas (SMSA's), were called self-representing (SR) because the sample from the sample area represented just that PSU. Each of the other 220 strata consisted of a group of PSU's and were referred to as nonself-representing (NSR), since the sample of housing units from the sample PSU in a stratum represented the other PSU's in the stratum as well.

One PSU was selected from each NSR stratum with probability proportionate to the 1970 census population to the PSU. (This resulted in 220 NSR sample PSU's.) In addition, the NSR strata were grouped into 110 pairs and 1 stratum was picked at random from each pair. From this stratum, an additional PSU was selected independently of the other PSU selected from this stratum. Since the two PSU's were independently selected, it was possible for the same PSU to be selected twice. This occurred in 25 instances, producing an additional 85 NSR sample PSU's, thus giving a grand total of 461 PSU's.

In 1974, it was decided to increase the reliability of the AHS estimates of rural housing characteristics by doubling the number of sample housing units from rural areas. This was accomplished by activating the reserve sample, selected in the original sampling operations in 1973, from rural areas only. For the reserve sample selected in census address and new construction frames, the other half of each rural cluster (an expected two housing units) was activated in 1974, if the cluster was rural. This supplementation increased the overall probability of selection for sample housing units in rural areas to about 1 in 683; whereas, the overall probability of selection for sample housing units in urban areas remained at about 1 in 1,366.

For the 1978 AHS, approximately 77,900 sample housing units were identified in the sample areas. Of this number, about 5,900 sample units were visited, but were found to be ineligible for interview for AHS in terms of collecting information relevant to the 1978 housing inventory. Another 4,400 units were eligible for interview, but were classified as "noninterview" for various reasons.

At each interviewed household, a supplemental set of agricultural screening questions was asked of all individuals enumerated in the 1978 AHS. These screening questions were comparable to screening questions asked of households in the CAAS. All identified potential farm operations were matched to the 1978 Census of Agriculture mail list. Non-match and doubtful match cases were mailed an evaluation report form to obtain the basic agricultural characteristics

of the operation and additional information for matching purposes.

Totals at the regional level were estimated for farm counts by major characteristics. Individual farm characteristics are expanded by the reciprocal of the probability of selection. Generally, the expansion factor was 683 in rural areas and 1,366 in urban areas. Estimates were made for total farms included farms, missed farms, and overcounted farms.

Sampling variance for major data items was estimated at the regional and national levels. Estimates of sampling reliability were made separately for the NSR and SR strata. In NSR strata, the 220 strata were collapsed into 110 strata. A third sample PSU was randomly selected from the two PSU's in each strata. The three PSU's were used to estimate variances in NSR strata. In SR strata, variances were estimated using a balanced half-sample replication estimator using all possible samples to produce estimate variances.

PES—The PES was a subsample of the 1978 CAAS. PES estimates are based on data collected in December 1978 through January 1979 following the completion of the CAAS enumeration.

The sample for the 1978 CAAS was selected from rural areas (areas in the 1970 Census of Population and Housing classified as having less than 2,500 inhabitants). A sample was selected independently from each State in the conterminous United States. A total of 6,391 sample areas were selected. Areas were selected separately from one of six strata defined by expected farm density. Data collection resulted in approximately 560,000 housing units screened and 92,000 potential farm operations identified.

The PES used a 1 in 30 subsample of the CAAS segments. The subsample was selected independently from each strata without consideration for State. Some States and groups of States had no subsample areas selected in a given strata. The PES sample was a stratified sample with an unequal probability sample within a strata. A total of 212 subsample areas were selected with approximately 16,000 households screened and approximately 3,500 potential farm operations identified.

All identified potential farm operations were matched to CAAS and to the 1978 Census of Agriculture mail list. Nonmatch and doubtful match cases were mailed an evaluation report form to obtain the basic agricultural characteristics of the operation and additional information for matching purposes.

Identified farms were weighted by the reciprocal of the probability of selection,

$$\text{weight} = 1978 \text{ CAAS weight} \times 30.$$

Sampling error is estimated by strata within regions. To estimate sampling error, all subsample areas in a given strata of a region were collapsed into one strata. Within a collapsed stratum of a region, sampling errors were estimated assuming unequal probability random sampling with replacement. The sampling variance of strata totals were summed to estimate the sampling variance of a regional total. Regional variances were added to calculate estimates for the national total.

