

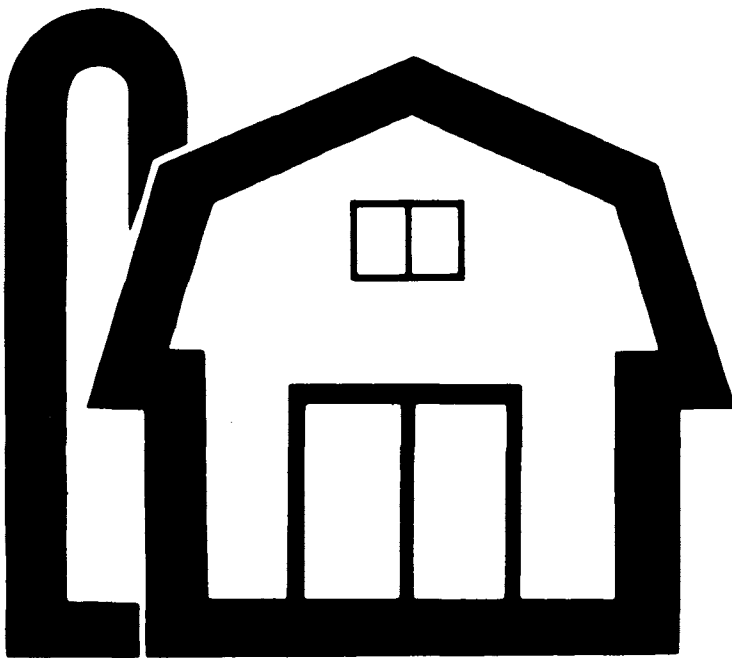
1982 Census of Agriculture

AC82-SS-2

Volume 2
SUBJECT SERIES

Part 2

Coverage Evaluation



The publications
from the 1982 Economic and
Agriculture Censuses are dedicated
to the memory of Shirley Kallek,
Associate Director for Economic Fields.
During her career at the Bureau of the
Census (1955 to 1983), she continually
directed efforts to improve
the timeliness and accuracy of
economic statistics.

1982 Census of Agriculture

AC82-SS-2

Volume 2
SUBJECT SERIES

Part 2

Coverage Evaluation

Issued April 1985



U.S. Department of Commerce
Malcolm Baldrige, Secretary
Clarence J. Brown, Deputy Secretary
Sidney Jones, Under Secretary for
Economic Affairs

BUREAU OF THE CENSUS
John G. Keane,
Director



BUREAU OF THE CENSUS
John G. Keane, Director
C.L. Kincannon, Deputy Director
Charles A. Waite, Associate Director
for Economic Fields
John H. Berry, Assistant Director for
Economic and Agriculture Censuses

AGRICULTURE DIVISION
Charles P. Pautler, Jr., Chief

ACKNOWLEDGMENTS—Many persons contributed to the coverage evaluation program for the 1982 Census of Agriculture. First and foremost are the many farm and ranch operators, members of their families, and employees who furnished the information. Only through their cooperation was it possible to collect and publish the data in this report. This program was part of the Census Bureau's overall program of evaluation for the 1982 Economic and Agriculture Censuses. It was coordinated by the 1982 Censuses Evaluation and Research Task Force headed by **Kirk Wolter**, Chief, Statistical Research Division.

Primary direction was by **Shirley Kallek**, Associate Director for Economic Fields (to May 1983) and **Charles A. Waite**, her successor, and **Michael G. Farrell**, Assistant Director for Economic and Agriculture Censuses (to August 1984), and **John H. Berry**, his successor. Technical direction and guidance were provided by **Barbara A. Bailar**, Associate Director for Statistical Standards and Methodology.

This program was developed in the Agriculture Division under the direction of **John H. Berry**, Chief (to August 1984) and **Charles P. Pautler, Jr.**, Chief, his successor.

The overall program was supervised by **D. Dean Prochaska**, Chief, Program Research and Development Branch. Significant technical and administrative contributions were made by **Jane Dea Sandusky**, **William C. Davie**, and **Emily Burton**. Sample design and estimation were provided by **David D. Chapman** and **Nicholas S. Alberti**, Agriculture Division, and **David W. Chapman** and **Carma Hogue**, Statistical Research Division.

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; Field—data collection and followup; and Publication Services—publication planning and design, editorial review, composition, and printing procurement.

Library of Congress Cataloging in Publication Data
(Revised for volume 2)

Main entry under title:

1982 census of agriculture.

Includes indexes.

Supt. of Docs. no. C 3.31/4:982/v.2/pt.1

1. Agricultural—United States—Statistics.

I. United States. Bureau of the Census.

HD1769.A14 1984 338.1'0973

83-600308

For sale by Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402.

If you have any questions concerning the statistics in this report, call (301) 763-5230.

CONTENTS

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

CHARTS

1. Estimate of Census Farm Coverage by Coverage Component and Region	XVI
2. Estimate of Net Census Farm Coverage by Value of Sales and Region	XVI

TABLES

1. Farms by Selected Characteristics and Components of Coverage	1
2. Selected Items for Estimated Undercounted Farms	4
3. Land in Farms by Sales Group and Components of Coverage	6
4. Value of Agricultural Products Sold by Sales Group and Components of Coverage	7

APPENDIXES

A. Report Form and Listing Sheet	A-1
B. References	B-1

Publication Program	Inside back cover
-------------------------------	-------------------

INTRODUCTION

PURPOSE OF COVERAGE EVALUATION	IV
CENSUS AUTHORITY	IV
FARM DEFINITION	IV

directs that the census be taken for 1978, 1982, and every fifth year thereafter.

PURPOSE OF COVERAGE EVALUATION

The Bureau of the Census seeks to measure the accuracy and completeness of farm counts and selected data items for each census of agriculture through a coverage evaluation program. This program provides an independent check on the census results. Also, the program aids in identifying problem areas associated with coverage errors as a basis for improving the census mail list, data collection, and data processing. The results from this program are an important means of informing the users of any known deficiencies which might affect their interpretation and use of the data.

CENSUS AUTHORITY

The census of agriculture is required by law under title 13, United States Code, sections 142(a) and 191, which

FARM DEFINITION

Since 1850, when minimum criteria defining a farm for census purposes were first established, the farm definition has been changed nine times. The current definition, first used for the 1974 final reports, is any place from which \$1,000 or more of agricultural products were sold or normally would 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 normally would 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 normally would have been sold during the census year. A place not having sufficient sales to qualify as a farm can qualify on potential sales based on the inventory and production of crops and/or livestock.

GENERAL EXPLANATION

1982 CENSUS OF AGRICULTURE	V
1982 COVERAGE EVALUATION PROGRAM	VI
RESULTS	X

1982 CENSUS OF AGRICULTURE

A brief summary is provided below as an introduction to the census. A more detailed description of the procedures is in volume 1, appendix A.

History

The 1982 Census of Agriculture was the 22d 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. This change in reference years provided for joint processing operations and more data comparability among the various censuses.

Data Collection

Before 1969, the census of agriculture was based on a canvass of rural areas by enumerators and personal interviews of farm operators. Beginning with 1969, censuses have been conducted primarily by mail. The 1978 census was the only census to include a personal interview of all households in a sample of area segments to supplement the mailout/mailback enumeration. The mailout/mailback enumeration, plus the area sample, was used to improve completeness of coverage for U.S., regional, and State level statistics. The area sample did not provide county estimates. Due to budget reductions, an area sample was not part of the 1982 Census of Agriculture.

Mail List

The mail list was comprised of all individuals, businesses, and organizations readily identified in a cost-effective manner as being associated with agriculture. The preliminary census mail list was assembled from the records of the 1978 census; administrative records of various government agencies, primarily the Internal Revenue Service and the U.S. Department of Agriculture;

and agriculture-related organizations. Lists of large and specialized operations were obtained from trade associations and various State and Federal agencies. The total number of records from all sources on the preliminary list was about 19.0 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 not likely to be farms were included in the 1982 Farm and Ranch Identification Survey. These records appeared on one source list or selected combinations of lists which had yielded a low percentage of farm operators in the 1978 census. These addresses were mailed a short screening report form to identify their current status. As a result of the Farm and Ranch Survey, addresses with no agricultural operations were excluded and new tenants and successors were added. The final list contained approximately 3.7 million names and addresses.

Report Form

The report form contained questions about land ownership, land use, crops, livestock and poultry, market value of agricultural products sold, expenses, and operator characteristics. Regional versions of the report form, listing crops and livestock commonly produced in each region, were used. This enhanced the reporting of crop and livestock data and reduced respondent burden.

Mailing and Followup

Report forms were mailed in late December 1982. Nonrespondents were sent a reminder card and five followup requests 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 (those with expected sales of \$100,000 or more). The data collection effort achieved a final response rate of about 86 percent. The final nonrespondent farms were represented in the census by a nonresponse adjustment procedure and results are subject to sampling variability. A description of the nonresponse adjustment procedure is included in volume 1, appendix A, **Statistical Adjustments**.

Data Processing

Report forms with attached correspondence, remarks, or missing data were reviewed prior to keying the data onto magnetic tape. All report forms were then keyed and

a detailed item-by-item computer consistency edit of the data was performed. Telephone calls were made to resolve conflicting data, verify large entries, or provide missing information. The data items were tabulated and reviewed by county to correct any remaining problems.

1982 COVERAGE EVALUATION PROGRAM

History

Coverage evaluation studies have been conducted for each census of agriculture since 1945. Several procedural modifications resulting from coverage evaluation findings have been introduced into various censuses.

For the 1945 census and previous censuses, interviewers were given the farm definition and instructed to enumerate all places conforming to that definition. The 1945 coverage evaluation study showed that marginal farming operations were a large proportion of the undercounted (missed) farms. In the 1950 census, to improve the coverage of these marginal operations, interviewers were instructed to enumerate all places with specified agricultural activities and the farm definition was applied during processing.

The 1950 evaluation study found nonresident farm operators to be a large share of the undercounted farms. In the 1954 census, two new techniques were introduced to reduce undercoverage.

- a. Enumerators in selected counties drew the boundaries of each farm and each nonfarm tract on a township sketch.
- b. A listing book was used to record the location and identification of every residence and every agriculture operation in each enumeration district.

A feasibility study using the 1964 evaluation sample was conducted prior to the 1969 Census of Agriculture. The study indicated that at least equal and perhaps better coverage could be obtained with a mailout/mailback procedure. A discussion of this study may be found in the 1964 Coverage Evaluation publication. The mailout/mailback method of data collection was first implemented in the 1969 census.

Coverage evaluations for 1969 and 1974 indicated that the source lists acquired for data collection by mail did not provide adequate coverage of small farms with sales of agricultural products of less than \$2,500. In 1969 and 1974, 33 percent (± 1.3)¹ and 27 percent (± 1.2), respectively, of all small farms were undercounted at the U.S. level. To reduce the sizable undercount of small farms, the 1978 census was supplemented by the Census of Agriculture Area Sample (CAAS). This supplemental survey reduced the undercount of small farms to about 7 percent (± 1.5) at the U.S. level.

Objectives

Although the goal of each census is to count all farms, it cannot realistically be attained. Complexity of farm organizational arrangements, continuing changes in operational status, inadequacies of source lists, difficulty in communicating census definitions and concepts, and other factors can contribute to census error and incompleteness.

The 1982 Coverage Evaluation program was designed to provide estimates of various aspects of census coverage at the U.S. and regional levels. The objectives were:

- a. To provide measures of accuracy of census farm counts by a limited number of items, such as land in farms, value of agricultural products sold, and operator characteristics.
- b. To provide selected undercount estimates of land, value of sales, and major crop and livestock data on undercounted farms.

Sample Design and Methodology

The 1982 Coverage Evaluation program consisted of two parts—Area Segment Survey and Classification Error Study.

Area Segment Survey—This survey was designed to measure the number and characteristics of farms operated by persons living in rural areas (areas with a population of less than 2,500 inhabitants in the 1970 Census of Population and Housing) who were not on the census mail list.

Due to budget restraints, no survey was done to measure the urban farms not on the mail list for the 1982 Coverage Evaluation program. However, the coverage error model presented in **Estimation Procedure** includes an estimate of the portion of urban farms not on the mail list. The 1978 Coverage Evaluation program found that less than 1 percent of all farms were urban farms not on the mail list. (See appendix B.) Therefore, the lack of a direct measurement for this portion of urban farms probably has little effect on the estimates in the coverage evaluation.

This survey was based on a sample of 344 land area segments. A segment is an area of land with boundaries identifiable on a map and on the ground. Because of the time requirements and the high costs of mapping land area segments, a subsample of 344 segments was 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. The sample frame for the CAAS was constructed using maps and data from the 1970 Census of Population and Housing. The sample frame consisted of rural areas with a population of less than 2,500 inhabitants. The CAAS was a stratified one-

¹Numbers in parentheses are estimated standard errors. See **Accuracy of the Estimates**.

stage cluster sample with stratification by State and by 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 average number of farms per segment was 10 and varied by stratum from none in low density farm areas to 12 in high density farm areas.

In 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 in CAAS as not on the mail list: 0, 1, 2 or 3, and 4 or more. Once ordered within the 16 strata by farm density ratio and segment number, 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 systematically from each stratum with probabilities proportional to the measures of size. Measures of size were used in the selection procedure to provide a sample for the 1982 Area Segment Survey in which each segment in each of the 16 strata had approximately the same overall selection probability. The measures of size provided an adjustment for the variation of the 1978 segment selection probabilities within each of the 16 strata.

In designing the sample, consideration was given to cost, precision, the importance of regional estimates, and availability of maps. The sample of 344 area segments was chosen to provide an absolute standard error of approximately 2.0 percent for the estimated proportion of farms not on the mail list at the regional level. Strata sample sizes within each region were based on an approximate optimum allocation of the sample, with the constraint that at least two segments be allocated to each stratum. The segments used in the 1982 sample were the segments adjacent to each selected CAAS segment. Adjacent segments were selected to avoid potential bias from the previous census enumeration. The selection probability of a 1982 sample segment was equal to the selection probability of the adjacent CAAS segment. The overall selection probability for each segment chosen for the 1982 Area Segment Survey was the product of two factors: (a) the selection probability for the 1978 CAAS, and (b) 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.

The data collection procedures were uniform across all segments with extensive emphasis on completeness and accuracy. Experienced enumerators were selected and trained under the supervision of the Census Field Division regional offices. Beginning 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 enumerators asked the reference person screening questions to determine if any person in the household was associated with 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 report form (see appendix A) was completed for each agriculture operation in the household. This form was an abbreviated version of the census report form. It contained questions about alternate farm names and addresses used for the operation, farm size, crops, livestock, and various operator characteristics. If a household could not be contacted by an enumerator, attempts were made by telephone and mail to complete the evaluation report form. Enumeration was completed by May 1983, and all evaluation report forms were returned to the Washington, D.C. office for processing.

Classification Error Study—This study was designed to measure the number and characteristics of farms on the census mail list, but overcounted or misclassified as non-farms. Since the census mail list included farms operated by persons residing in both urban and rural areas, the Classification Error Study measured error for farms in both urban and rural areas.

Classification error contributes to coverage error in the census of agriculture. Coverage evaluation results from 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 were nonfarms incorrectly classified as farms or farms with more than one report in the census and were, therefore, overcounted. Classification error may result from misinterpretation of census definitions and instructions, incomplete reporting by respondents, and errors in census processing.

The sample for the Classification Error Study was a multistage sample selected from the census mail list of 3.7 million names and addresses. Addresses in Alaska and Hawaii were excluded because of limited evaluation funds. Also, 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 followup, telephone followup, and report form review to ensure the accuracy and completeness of their data. The first stage was the selection of a systematic sample of about 4,700 names and addresses from the census mail list with a sampling rate that varied by census geographic region: 1 in 187 in the Northeast, 1 in 1,250 in the Midwest and South, and 1 in 375 in the West. These rates resulted in a sample with approximately equal numbers of names and addresses from each region. With consideration for cost and precision, this sample was of sufficient size for acceptable regional estimates of classification error.

The census report forms mailed to the 4,700 sample addresses were identified with a special identification symbol on the report form name and address label. The symbol was used only to separate the sample report forms for photocopying after they were returned. The respondents and processing staff were unaware of the special symbol. After photocopying, the forms were returned to regular processing. The photocopies of the sample report forms were reviewed and classified into four groups: 2,700 farms, 1,400 nonfarms, 500 nonrespondents, and 100 postmaster returns (PMR's) undeliverable by the post office.

The second selection stage consisted of all nonfarm cases, all PMR's, and a 1 in 2 subsample of the farm cases. The sampled farm cases were matched to the census mail list for selection of all potential duplicates for reenumeration since these cases were more likely to be overcounted. Nonduplicated farm cases were systematically subsampled with rates varying by census geographic region: 1 in 5 in the Northeast, Midwest, and South; and 1 in 7 in the West. These rates provided a level of accuracy comparable to the Area Segment Survey. The potential duplicates and the subsample of nonduplicates resulted in about 300 farm cases chosen for reenumeration.

Data for the Classification Error Study were collected primarily by telephone interviews. For the telephone reenumeration, experienced Census Bureau personnel conducted intense interviews using the evaluation report form. If a household could not be contacted by telephone or if a telephone number could not be obtained, attempts were made to obtain the information by mail.

The evaluation report forms were compared to the census report forms to identify errors. The types of census errors identified were: (a) farms misclassified as nonfarms; (b) farms that were PMR's in the census (*not classified* in the census); (c) nonfarms incorrectly classified as farms; and (d) more than one report form for the same farm, where: (a) and (b) were *misclassified* farms and (c) and (d) were *overcounted* farms.

Because the nonresponse adjustment procedure in the census does not uniquely relate values to individual nonrespondent addresses, the Classification Error Study could not be used to measure error in classification of nonrespondents. No further investigation was done for the 500 census nonrespondents in the study.

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. Forms that could not be classified as farm or nonfarm were considered to be unclassified cases. (See **Nonsampling Error**.)

The classified evaluation report forms were then matched to the census mail list using information obtained from the interviews, in particular, alternate names and/or addresses. Area segment farms identified as farms on the census mail list were classified as *matched* farms; non-matched area segment farms were classified as farms *not on the mail list*. *Misclassified* and *overcounted* farms were measured by the Classification Error Study.

Following the matching, a final review was completed. Coverage classification codes were assigned for types of census errors. The data were then keyed, edited, reviewed for accuracy and consistency, and tabulated providing sample estimates and variances for publication.

Estimation Procedure

The coverage evaluation provides an estimate of the undercount and the overcount. The undercount component is available for farm counts and totals of various farm characteristics, and the overcount component is available only for farm counts. Any total (T) for some characteristic of all farms in the United States can be represented as the census published number (C) for that characteristic plus the undercount (U) for that characteristic minus the overcount (OV) for that characteristic; i.e.:

Equation 1:

$$T = C + U - OV$$

The undercount (U) can be split into a component consisting of farms not on the census mail list (NML) and a component consisting of farms on the census mail list that were misclassified as nonfarms (MCF), substituting into equation (1):

Equation 2:

$$T = C + NML + MCF - OV$$

The estimates of the overcount ($\hat{OV}$) and of totals of characteristics for farms on the census mail list misclassified as nonfarms (MCF) are unbiased sample estimates from the Classification Error Study. An estimation procedure for farms not on the mail list was chosen that would account for the absence of coverage measures of urban farms not on the mail list. This procedure assumes that urban farms and rural farms not on the mail list have similar characteristics, an assumption supported by information from the 1978 Coverage Evaluation program.

The estimation procedure for farms not on the mail list is based on a coverage error model that has the following assumptions: a) both the census (List A), which is observable, and the universe of the Area Segment Survey (List B), which is not observable, attempt to accurately enumerate the complete universe of farms and that farms reported on either list are true farms; b) the event of being missed by the census is independent of being missed in the survey; and c) the probability of being missed by either the census or the survey is the same for all farms within a given size category. Since both lists may be incomplete, each farm in the universe can be placed into one of the cells in the following matrix.

Census farms (List A)	Area Segment Survey farms (List B)	
	In the survey universe	Out of the survey universe
On the mail list	N_{11}	N_{12}
Not on the mail list	N_{21}	N_{22}

N_{11} = the number of farms on the mail list and in the Area Segment Survey universe.

N_{12} = the number of farms on the mail list but not in the Area Segment Survey universe.

N_{21} = the number of farms not on the mail list but in the Area Segment Survey universe.

N_{22} = the number of farms not on the mail list and not in the Area Segment Survey universe.

$N_c = N_{11} + N_{12}$, (the number of farms on the mail list, i.e., the census total farm count).

The estimate of N_{12} is:

Equation 3:

$$\hat{N}_{12} = N_c - \hat{N}_{11},$$

and the estimate of N_{22} is:

Equation 4:

$$\hat{N}_{22} = \frac{\hat{N}_{21} \hat{N}_{12}}{\hat{N}_{11}}$$

where $\hat{N}_{11}$ and $\hat{N}_{21}$ are design based estimates resulting from the match of Area Segment Survey sample farms to farms in the census. The properties of this estimate and its derivation are discussed by Wolter, (1983). (See appendix B.)

The estimate of the total number of farms not on the mail list is $\hat{N}_{21} + \hat{N}_{22}$. The component of the undercount of some characteristic for farms not on the census mail list (NML) is estimated by:

Equation 5:

$$\begin{aligned}\widehat{NML} &= (\hat{N}_{21} + \hat{N}_{22}) (\hat{S}/\hat{N}_{21}) \\ &= \hat{S} (N_c/\hat{N}_{11})\end{aligned}$$

where $\widehat{NML}$ is the estimate of some characteristic of farms not on the mail list and $\hat{S}$ is the unbiased sample estimate of the total of some characteristic for farms not on the mail list but in the Area Segment Survey universe. As an estimate of the total number of farms not on the census mail list, $\widehat{NML}$ becomes $\hat{N}_{21} (N_c/\hat{N}_{11})$. The estimated total $\hat{T}$ for some characteristic is obtained by rewriting T in terms of the estimated components:

Equation 6:

$$\hat{T} = C + \hat{S} (N_c/\hat{N}_{11}) + \widehat{MCF} - \widehat{OV}$$

The estimates $\hat{S}$ and $\hat{N}_{11}$ were computed independently for farms with sales of less than \$2,500 and for farms with sales of \$2,500 or more within each region and summed to produce the region total. The estimates of U.S. totals are the sum of the regional level estimates.

The estimates of the proportions for components of coverage error are of the form:

a. Farms not on the census mail list (percent) =

$$\frac{\widehat{NML}}{\hat{T}} (100).$$

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

c. Undercounted farms (percent) = $\frac{\hat{U}}{\hat{T}} (100).$

d. Overcounted farms (percent) = $\frac{\widehat{OV}}{\hat{T}} (100).$

e. Net coverage (percent) = $\frac{\hat{U} - \widehat{OV}}{\hat{T}} (100).$

Variance Estimation

Estimates in this evaluation study are based on independent surveys—the Area Segment Survey and the Classification Error Study. Estimates of the totals and their variances are the sum of estimates from these two separate surveys. The estimation of sampling variances are discussed separately for each survey.

Area Segment Survey—As described previously in **Sample Design and Methodology**, the sample for the Area Segment Survey was a one-stage stratified cluster subsample of area segments selected with unequal probabilities from 1978 CAAS sample. Each farm identified by the survey was weighted by the reciprocal of its probability of selection which was equal to the probability of selection of the segment in which it was located. Sampling errors for the unbiased estimates of totals for farm characteristics (i.e., $\hat{N}_{11}$ —the estimated number of farms on the census mail list and in the Area Segment Survey universe, and $\hat{S}$ —the estimated characteristic totals of farms not on the mail list but in the Area Segment Survey universe) were estimated assuming unequal probability sampling with replacement.

The sampling errors for the final survey estimates of totals for characteristics of farms not on the mail list ($\widehat{NML}$) were estimated by using a Taylor series expansion to approximate the nonlinear estimator $\widehat{NML}$ by a linear function of $\hat{N}_{11}$ and $\hat{S}$, the variances of these component estimates, and the covariance between these component estimates. Regional variances for estimated totals are the sum of variances for strata within each region. The U.S. level variances are the sum of regional variances.

Classification Error Study—As indicated in the description of the sample selection for the Classification Error Study, there are two stages of sample selection. All cases selected at the first stage were divided into two groups based on preliminary census mail return status: a) non-farms and PMR's, and b) farms. All cases in the first group were selected for reenumeration and a subsample of cases in the second group was selected for reenumeration. Because census returns were edited after these two groups were formed, the status of some returns in both groups changed. Thus both undercounted and overcounted farms were counted in each of the two groups.

Each case in both groups was weighted by the reciprocal of their probability of selection. Estimates of classification error are the sum of separate estimates for each of the two groups. Variances within each group were estimated assuming a simple random sample was chosen. Sampling errors of estimated undercount due to misclassification and overcount are the sum of the estimated variances for estimates within each group plus the estimated covariances between these separate group estimates.

RESULTS

Estimates of Census Coverage

Estimates of census coverage were made only at regional and national levels since evaluation samples were too small to provide reliable estimates at State or county levels. Estimates are based on a combination of the Area Segment Survey and Classification Error Study samples.

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 undercounted farms or the proportion of undercounted farms alone are not a complete indication of the quality of the census data. Consideration of economic characteristics along with the farm counts may be a better indication of census quality and may have a greater impact on the user's needs. For example, while the estimated net undercounted farm rate was 9.1 percent for the United States, the undercounted farms accounted for only 1.8 percent (± 0.5) of the estimated value of agricultural products sold and only 2.1 percent (± 0.3) of the estimated land in farms.

Regional estimates are presented in tables 1 through 4 to provide some indication of census coverage below the national level. Because of the relatively high sampling error, especially in the overcount component, caution should be observed when drawing conclusions based upon comparisons of regional estimates within and between tables. In addition, coverage for States or counties within a region may be highly variable.

Table 1 presents the number of farms by sales group, standard industrial classification, size in acres, and operator characteristics by components of coverage.

Farms are classified as census farms, undercounted, and overcounted farms. Overcounted farms are part of the farms in the census. Table A presents selected U.S. values from table 1 as a percent of estimated totals.

Estimates of net census coverage indicate that 90.9 percent of the estimated total farms were in the 1982 census for the conterminous United States. Approximately 13.7 percent of estimated total farms were undercounted and approximately 4.6 percent were overcounted resulting in a net undercounted rate of about 9.1 percent for data at the U.S. level. The gross undercounted rate was 17.6 percent (± 0.5) in 1969, 12.7 percent (± 0.4) in 1974, and 4.4 percent (± 0.5) in 1978. The considerable improvement for the gross undercount rate in 1978 resulted primarily from the inclusion of the area sample. Because no area sample was conducted in 1982, the gross undercount rate increased substantially. For comparisons of 1982 data with data from previous censuses, see Clark, (1984). (See appendix B.)

An estimated 71.4 percent of farms with value of agricultural products sold of less than \$2,500 were in the census. About 35.4 percent (± 2.7) of these farms were undercounted and 6.8 percent (± 2.8) were overcounted. The gross undercounted farm rate for this group was 33.3 percent (± 1.3) in 1969, 27.3 percent (± 1.2) in 1974, and 7.1 percent (± 1.5) in 1978.

For farms with value of agricultural products sold of \$2,500 or more, 99.5 percent were in the census. This group was derived from table 1 by combining the \$2,500 to \$9,999 and \$10,000 or more sales groups. About 4.2 percent (± 0.7) of the \$2,500 or more sales group were undercounted and 3.7 percent (± 1.4) were overcounted. The undercounted farm rate for this group was 6.5 percent (± 0.4) in 1969, 6.8 percent (± 0.3) in 1974, and 3.6 percent (± 0.2) in 1978. 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.

Census coverage in the Midwest Region was substantially more complete for all farms than in the Northeast, South, and West. This is due primarily to the higher proportion of larger farms in the Midwest which are more likely to be included on the sources for the mail list.

The estimated number of overcounted farms was 113,623 or 4.6 percent. About four out of five overcounted cases were nonfarms incorrectly classified as farms. The remainder were farms with more than one census report (reports duplicated for a single farm or multiple reports for parts of a single farm).

The estimated number of undercounted farms was 336,498 or 13.7 percent. About 3 of 4 undercounted farms were not on the mail list. The remainder were farms misclassified as nonfarms. While about 21 percent of the undercounted farms had value of agricultural products sold of \$2,500 or more, only about 5 percent were larger farms with sales of \$40,000 or more. Of the undercounted farms, about 83 percent had less than 100 acres,

Table A. Coverage Percents by Selected Characteristics and Components of Coverage

Characteristics	Estimated undercounted farms ¹		Estimated overcounted farms		Estimated net undercount ²	
	Percent	Relative standard error (percent)	Percent	Relative standard error (percent)	Percent	Relative standard error (percent)
Total	13.7	8.2	4.6	27.1	9.1	17.7
Farms by value of sales:						
Less than \$2,500	35.4	7.5	6.8	40.6	28.6	11.6
\$2,500 or more	4.2	17.1	3.7	37.0	0.5	(³)
\$2,500 to \$9,999	5.9	17.2	5.3	48.4	0.6	(³)
\$10,000 or more	3.4	28.4	2.9	55.6	0.5	(³)
Farms by standard industrial classification:						
Crops (01)	7.0	12.4	6.9	38.3	0.1	(³)
Livestock (02)	18.5	9.2	3.0	35.9	15.5	12.6
Farms by size:						
1 to 99 acres	23.7	8.4	6.3	33.3	17.4	15.3
100 to 499 acres	5.2	15.0	4.0	52.1	1.2	(³)
500 acres or more	2.4	41.1	0.8	63.4	1.6	68.3
Farms by tenure of operator:						
Full owners	16.6	10.0	7.7	28.4	8.9	28.9
Part owners	7.7	15.2	0.2	80.6	7.5	15.9
Tenants	13.6	14.6	0.1	(³)	13.5	14.6
Farms by age of operator:						
Under 35 years	16.2	20.2	1.6	50.0	14.6	23.4
35 to 54 years	15.1	10.0	3.7	48.9	11.4	19.0
55 years and over	11.1	14.3	7.0	36.2	4.1	68.1
Farms by principal occupation of operator:						
Farming	7.3	16.8	1.6	26.3	5.7	22.8
Other	20.9	8.7	8.1	33.8	12.8	22.8

¹Referred to as missed farms in previous evaluation reports.²Undercounted farms minus overcounted farms.³Relative standard error is over 100 percent.

and about 3 percent had 500 acres or more. Of the undercounted farms, about 72 percent were operated by full owners, about 16 percent by part owners, and about 12 percent by tenants. About 3 of 4 undercounted farms were livestock farms and 1 of 4 were crop farms.

The estimated relative standard error for the estimated total farms in the United States is 1.8 percent, and ranges from 2.7 percent to 3.7 percent in the regions. See Example 1 in **Accuracy of the Estimates**. The estimated relative standard error for the number of undercounted farms is 9.4 percent at the U.S. level and ranges from 12.9 percent to 29.4 percent in the regions.

Table 2 presents estimates of selected characteristics of farms not on the mail list and misclassified farms. Also, the estimated total undercount and their relative standard errors for these characteristics are given. These estimates do not represent total error in the census for these characteristics because detailed data for overcounted farms could not be derived from the coverage evaluation

sample and reporting error on correctly counted farms was not measured.

The estimated number of farms not on the mail list was 259,944 and the estimated number misclassified was 76,554 for an estimated total of 336,498 undercounted farms. The estimated undercounted acres were approximately 20.4 million with the average size undercounted farm having 60 acres compared to 440 acres for the average census farm.

Table B presents estimates of total undercount for selected crop and livestock items from table 2 and ratios of estimated undercount to estimated total by item. The estimated total does not include an estimate of overcounted farms. While the estimates of undercount probably understate the total error, the undercounted farm estimates for these items are likely to contribute substantially more to total error than other components such as overreporting and underreporting of specific detailed data for farms in the census.

Table B. Selected Items for Census Farms and Undercounted Farms

Item		Estimated farms ¹	Census published farms	Estimated undercounted farms ²	Ratio of undercounted farms to estimated farms (percent)
Corn for grain	farms. .	736 601	714 687	21 914	3.0
	acres. .	70 874 981	69 769 530	1 105 451	1.6
Sorghum for grain	farms. .	(³)	93,587	(³)	(³)
	acres. .	(³)	12,665,856	(³)	(³)
Wheat	farms. .	450,861	445,736	5,125	1.1
	acres. .	71,481,200	70,864,672	616,528	0.9
Soybeans	farms. .	524,147	510,958	13,189	2.5
	acres. .	65,706,137	64,791,074	915,063	1.4
Hay	farms. .	1,131,767	1,049,865	81,902	7.2
	acres. .	58,118,684	56,506,322	1,612,362	2.8
Tobacco	farms. .	189,213	179,103	10,110	5.3
	acres. .	945,274	931,183	14,091	1.5
Cattle and calves inventory	farms. .	1,577,855	1,352,916	224,939	14.3
	number. .	107,268,350	103,655,183	3,613,167	3.4
Hogs and pigs inventory	farms. .	406,743	329,031	77,712	19.1
	number. .	57,616,393	55,169,987	2,446,406	4.2
Hens and pullets inventory	farms. .	295,212	212,149	83,063	28.1
	number. .	310,488,880	308,978,702	1,510,178	0.5

Note: Detail may not add to total due to rounding.

¹Census published farms plus estimated undercounted farms.

²Referred to as missed farms in previous evaluation reports.

³No coverage error observed.

Table 3 presents estimates of the land in farms by sales group and by components of undercoverage. The estimated total land in farms was derived as the sum of the census published acres and the coverage estimate of undercounted acres. On this basis, it was estimated that 97.9 percent of the land in farms in the United States was in the census, with an estimated relative standard error of 0.3 percent. Undercounted farms accounted for 1.0 percent of the estimated total acres for farms with sales of \$2,500 or more, and 20.8 percent for farms with sales of less than \$2,500. The estimates for land in farms were based only on sample estimates for undercounted farms and do not represent total error, because reporting error was not measured for either farms in the census or overcounted farms.

Table 4 presents the estimates for the value of agricultural products sold by sales group and by components of undercoverage. The estimated total value of agricultural products sold was derived as the sum of the census published value and the coverage estimate of undercounted value. On this basis, it was estimated that 98.2 percent of the value of agricultural products sold in the United States was in the census, with an estimated relative standard error of 0.5 percent. Undercounted farms accounted for 1.7 percent of the estimated value of agricultural products sold for farms with sales of \$2,500 or more, and 18.0 percent for farms with sales of less than \$2,500. The estimates for value of agricultural products sold were based only on sample estimates for under-

counted farms and do not represent total error, because reporting error was not measured for either farms in the census or overcounted farms.

Accuracy of the Estimates

Two types of errors are possible in estimates based on a sample—sampling and nonsampling. Also, there may be a statistical bias in an estimator, but generally this bias is small and decreases with increasing sample size. Sampling error occurs because observations are made only on a sample and not on 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 (nonrespondents, refusals, incomplete report forms), response error, misinterpretation of definitions and concepts, 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. Extensive efforts were made to minimize coding and processing errors through the use of quality control and other verification measures.

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 generally 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 all 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.

Estimates of sampling variability are expressed as relative standard errors in tables 1 through 4. The estimated relative standard error (percent) for a statistic is derived by dividing the estimated standard error for the statistic by the statistic and multiplying by 100. Estimated relative standard errors are high for some regional estimates and the data should be used with caution.

The sample estimates and the estimates of relative standard errors 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 in table 1 is 2,457,017 with an estimated standard error of 43,008 and an estimated relative standard error of approximately 1.8 percent (i.e., 43,008 divided by 2,457,017 times 100). The chances are about 2 out of 3 (67 percent) that complete coverage using the same methods would yield between 2,414,009 and 2,500,025 farms. As calculated, the standard error does not measure the effect of nonsampling errors.

The relative standard errors for the components of coverage error in tables 1 through 4 may also be used to calculate approximate standard errors of the estimated net coverage error (i.e., undercount minus overcount), the estimated universe totals, and ratios of estimated components of coverage error to the estimated universe totals.

The standard error of the estimated net coverage error for each characteristic in table 1 may be estimated by the formula:

Equation 7:

$$\hat{\sigma}(\hat{E}) = \sqrt{\hat{U}^2 \hat{V}_1^2 + \hat{OV}^2 \hat{V}_2^2},$$

where $\hat{E}$ is the estimated net coverage error ($\hat{U} - \hat{OV}$), $\hat{U}$ is the estimated undercount, $\hat{OV}$ is the estimated overcount, and $\hat{V}_1$ and $\hat{V}_2$ are the estimated relative standard errors of $\hat{U}$ and $\hat{OV}$, respectively. The standard error for each estimated total ($\hat{T}$) in table 1 is equal to the standard error of the estimated net coverage error. In tables 3 and

4 the standard error of the estimated total ($\hat{T}$) is equal to the standard error of the estimated undercount since the overcount was not measured for the characteristics appearing in these two tables.

Example 1—To compute the standard error of the net coverage error for the total number of farms in the Northeast Region, apply the formula as follows:

$$\begin{aligned}\hat{U} &= 36,256 \\ \hat{OV} &= 5,998 \\ \hat{V}_1 &= .145 \\ \hat{V}_2 &= .457 \\ \hat{E} &= \hat{U} - \hat{OV} = 30,258 \\ \hat{\sigma}(\hat{E}) &= \sqrt{(36,256)^2 (.145)^2 + (5,998)^2 (.457)^2} \\ &= 5,928.8\end{aligned}$$

The estimated relative percent standard error of $\hat{E}$ is equal to:

$$\begin{aligned}\hat{V}(\hat{E}) &= \frac{\hat{\sigma}(\hat{E})}{\hat{E}} (100) \\ &= \frac{5,928.8}{30,258} (100) \\ &= .196 (100) \\ &= 19.6 \text{ percent}\end{aligned}$$

The standard errors of the estimated total (as computed in Example 1) and the standard errors of each estimated coverage error component (i.e., undercount, overcount, or net coverage error) for each characteristic may be used to approximate the standard error of the ratio $\hat{R} = \hat{X}/\hat{T}$, where $\hat{X}$ is the estimated coverage error component and $\hat{T}$ is the estimated total. The standard error of $\hat{R}$ can be approximated by the formula:

Equation 8:

$$\hat{\sigma}(\hat{R}) = \sqrt{\hat{T}^{-2} \hat{\sigma}^2(\hat{X}) (1 - 2\hat{R} + \hat{R}^2)},$$

where $\hat{\sigma}(\hat{X})$ is the estimated standard error of the estimated coverage error ($\hat{X}$).

Example 2—To compute the standard error of the ratio of the estimated net coverage error for farms to the estimated total farms in the Northeast Region, apply the formula as follows:

$$\begin{aligned}\hat{U} &= 36,256 \\ \hat{OV} &= 5,998 \\ \hat{E} &= \hat{U} - \hat{OV} = 30,258 \\ \hat{T} &= 162,101 \\ \hat{R} &= \hat{E}/\hat{T} = .187 \\ \hat{\sigma}(\hat{E}) &= 5,928.8 \text{ (from Example 1)} \\ \hat{\sigma}(\hat{R}) &= \sqrt{(162,101)^{-2} (5,928.8)^2 (1 - .374 + (.187)^2)} \\ &= .030\end{aligned}$$

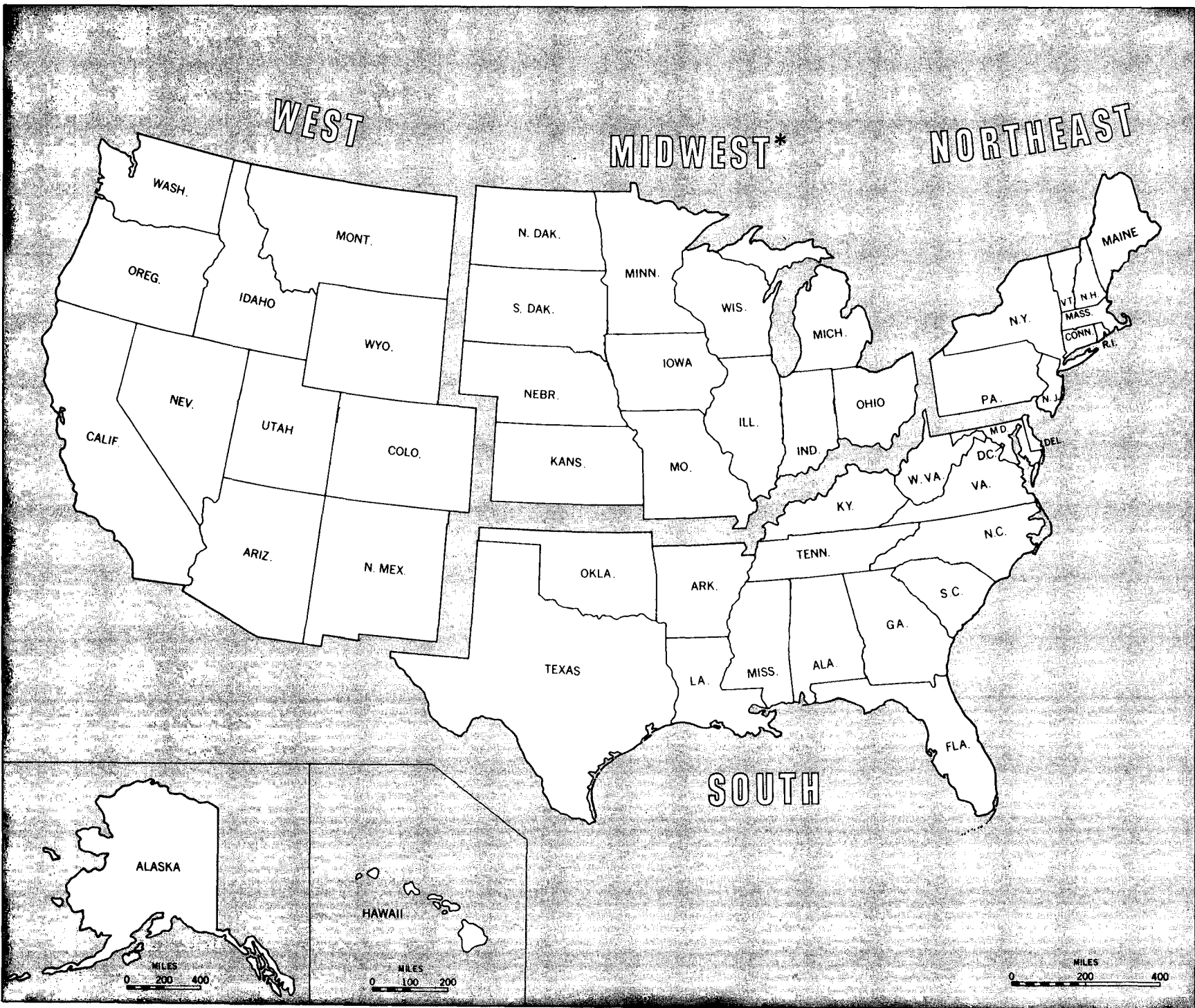
Nonsampling error—One source of possible nonsampling error in the coverage evaluation was the failure to classify about 4 percent of the evaluation report forms. These unclassified cases resulted from households that could not be contacted after a minimum of three attempts, households that refused to be interviewed, and households that

did not give the minimum information required for classification. If the correct classification could have been determined, the unclassified group most likely would have been spread throughout all coverage classification codes. If errors were present in the unclassified group, it is likely that they would be concentrated more heavily in the misclassified and overcounted farm components since the majority of the unclassified cases came from the Classification Error Study. No separate adjustment was made to the coverage estimates for the unclassified cases, thus resulting in a small downward bias in the estimates for

misclassified and overcounted farms and a slight downward bias for the estimated totals.

Another source of possible nonsampling error is related to the nonresponse adjustment procedure. This adjustment procedure makes the assumption that the respondents and the nonrespondents have similar statistical properties. The nonresponse adjustment represented about 10 percent of the farms and about 4 percent of the value of agricultural products sold in the 1982 census. Further explanation of the nonresponse adjustment procedure may be found in volume 1, appendix A.

REGIONS OF THE UNITED STATES



* The Midwest Region was designated as the North Central Region until June 1984.

U.S. DEPARTMENT OF COMMERCE
Bureau of the Census