1964 UNITED STATES CENSUS OF AGRICULTURE

FARMS
FARM CHARACTERISTICS
LIVESTOCK AND PRODUCTS
CROPS
FRUITS
VALUES

VOLUME III PART 7

Evaluation Surveys

CHAPTER 1

Coverage Check

U.S. DEPARTMENT OF COMMERCE **Maurice H. Stans,** Secretary
Rocco C. Siciliano, Under Secretary
BUREAU OF THE CENSUS **George Hay Brown,** Director

Reports for the 1964 Census of Agriculture

Preliminary reports

These reports, issued for each county and State, contain totals for farms, farm acreage, farm operators, land in farms classified by use, land-use practices, equipment and facilities, expenditures, use of agricultural chemicals, poultry, livestock, poultry and livestock products, and crops harvested, including fruits and nuts, nursery and greenhouse products, and forest products.

VOLUME I. STATE AND COUNTY STATISTICS

A separate part was issued for each State, Puerto Rico, Guam, and the Virgin Islands.

VOLUME II. GENERAL REPORT

Statistics by subject are presented, with totals for the U.S. regions, geographic divisions, and States.

VOLUME III. SPECIAL REPORTS

Reports present data for supplementary surveys for farm workers, hired farm workers, for the 1965 Sample Survey of Agriculture, farm debt, and reports on procedures and evaluation.

Issued October 1969

Suggested Citation

U.S. Bureau of the Census. Census of Agriculture, 1964

Volume III Special Reports
Part 7 Evaluation Surveys
Chapter I The Coverage Check

U.S. Government Printing Office, Washington, D.C. 1969

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

BUREAU OF THE CENSUS

George Hay Brown, Director
Robert F. Drury, Deputy Director
Joseph F. Daly, Associate Director for Research and Development
Conrad Taeuber, Associate Director for Demographic Fields

AGRICULTURE DIVISION
J. Thomas Breen, Chief
STATISTICAL RESEARCH DIVISION
Thomas B. Jabine, Chief

PREFACE

This report describes the methodology and the results of a study conducted to evaluate coverage in the 1964 Census of Agriculture. The study was limited primarily to measuring the accuracy of census data on number of farms, land in farms, and cropland harvested. The report includes a review of the uses which have been made of results from previous coverage checks and presents a statement on implications for future censuses of agriculture.

The coverage check was only one of two studies designed to evaluate the quality of information collected in the 1964 Census of Agriculture. A report on the other study, called the panel evaluation survey, will be published as chapter 2 of volume III, part 7.

The procedures and sample design for the 1964 coverage check were developed by Roe Goodman with the assistance of Donald G. Larson, both of the Statistical Research Division. The collection of data in the first and second phases of the field enumeration was under the direction of Jefferson D. McPike, Chief, Field Division, assisted by George Klink, Assistant Chief.

The Statistical Laboratory of Iowa State University took part in the preparation of questionnaires and instruction materials and conducted the field work and certain phases of the processing for parts of the country, under a contract with the Bureau of the Census. This work was supervised by Professors Wayne Fuller, Norman Strand, and Harold Baker. Charles K. Graham, of the Statistical Laboratory, supervised and carried out the analysis of the two methodological studies described in section VIII, part B.

Later stages of the field work and processing were coordinated by Eugene Hixson of the Statistical Research Division. Detailed specifications for computer tabulations were prepared by Thomas B. Jabine, of the Office of the Associate Director for Research and Development, who also supervised the IRS match project described in section VIII, part A. Computer programs were prepared by Alfred Sands of the Statistical Research Division.

This report was prepared by Bryan J. Hargis of the Statistical Research Division. General supervision throughout the project was provided by William N. Hurwitz, formerly Chief of the Statistical Research Division, and Ray Hurley, formerly Chief of the Agriculture Division.

1964 UNITED STATES CENSUS OF AGRICULTURE

Volume III—Part 7—Chapter 1

CONTENTS

CONTENTS—Continued

Section I. Introduction and Summary

A. Purpose of Census Coverage Checks[1]

The Bureau of the Census attempts to measure the accuracy of the statistics it produces in all major censuses. An important source of such measurement for agricultural statistics has been the coverage checks conducted in connection with each quinquennial census of agriculture since 1945. The primary purposes of these studies have been--

1. To identify factors associated with coverage errors, as an aid in planning future censuses.
2. To inform users about errors in the data which might affect their interpretation and uses of the data.

B. Objectives

The specific objectives of the coverage checks have been--

1. To provide national and regional measures of the accuracy of the census farm counts.
2. To provide measures of the accuracy of census results for a limited number of items such as land in farms, cropland harvested, acreage of major crops, and number of farms classified by tenure and size.
3. To provide information on factors associated with measurement errors, with special attention to the characteristics of farms missed in the census.

C. Principal Operations

The same principal operations have been used in all of the coverage checks to date. They are as follows:

Step 1. An enumeration, using the most reliable techniques available and without reference to the census results, of farms associated with segments in a probability sample of land areas.

Step 2. Matching results obtained in step 1 against census materials in order to identify farms missed in the census and differences between census and coverage check information for farms included in the census.

Step 3. Mail, telephone, and field followups, as needed, to clarify and check the results obtained in steps 1 and 2 and to obtain additional data for missed farms.

Step 4. Final processing, tabulation, and analysis of the results.

D. Types of Measurement Error[2]

To define measurement errors, it is convenient to assume that there exists for each statistic obtained in a census a *desired* or *true value.* The error in a census statistic is simply the difference between the statistic and its true value.

It is useful to divide the measurement errors in a census conceptually into two components--*response variance and bias.* To do this, it is necessary to assume that the census is a repeatable process of measurement, i.e., that independent census enumerations could be carried out with some conditions, such as the form of the questionnaire and written instructions, held constant but with other conditions, such as the particular persons selected as enumerators, and the time of day a particular farm operator is interviewed, subject to random fluctuations.[3]

For a given statistic, such as total land in farms, each of these repetitions of the census would produce a different value of the statistic. These values have a mean or expected value. The difference between this mean and the true value is the *bias,* and the variability of the values of the statistic (resulting from repetitions of the census enumeration) about their mean is the *response variance.* The net resultant of these two components of error is the *mean square error,* which is a measure of the variability of the values of the statistic about the true value.

For national and regional totals of most U.S. agricultural statistics, the error due to response variance is probably insignificant in comparison to the bias. The response variance arises from factors which tend to average out through compensating errors when large numbers of enumerators and/or respondents are involved, whereas the bias, although it may differ considerably for different areas or different censuses, is essentially independent of the size of the population. For smaller areas, such as counties and townships, however, response variance may be a significant source of error.

The above definitions are appropriate for census items for which data are collected from all farms. If some items are collected on a sample basis, sampling variability must also be considered along with the other components of error.

E. Summary of Results

The results of the coverage check are presented in section VII. A summary of these results is given in the paragraphs which follow.

The coverage check estimate for number of farms was 3,559,000, compared to the census count of 3,158,000. The difference, or net error, of 401,000 was 11.3 percent of the coverage check total. For total land and cropland harvested, the estimated net errors relative to the coverage check estimates were smaller, being 6.1 and 6.0 percent, respectively.

---

[1] Coverage checks conducted prior to 1964 have generally been referred to in publications as "evaluation surveys." In this report we have adopted the term "coverage check" because there was a second evaluation survey, the panel evaluation survey, conducted in connection with the 1964 Census of Agriculture.

[2] For a more detailed and mathematical treatment of this subject, see appendix A, reference 2.

[3] In practice, of course, independent repetitions of a census cannot be realized; however, the model can reasonably approximate actual census conditions.

As in previous censuses, the relative net error in the farm count was greater for the smaller farms, whether size was measured in terms of total acres, acres of cropland harvested, or value of sales. The estimated net error for number of farms by value of sales was 3.1 percent of the coverage check estimate for farms with sales of $10,000 and over, 5.4 percent for farms with sales between $2,500 and $9,999, and 19.3 percent for farms with sales of less than $2,500. The estimated net coverage error for total value of sales was only 2.9 percent of the coverage check estimate.

For the most part, the levels of error observed in the 1964 Census of Agriculture did not differ greatly from those noted in the 1954 and 1959 censuses. An exception was the relative net error for number of farms with sales of less than $2,500, which was estimated at 19.3 percent in 1964 compared to 13.7 percent in 1959.

Differences in reporting total acres resulted in 15.4 percent of the correctly counted farms (those for which there was a census farm corresponding to the farm identified in the coverage check) being classified in different size classes in the census and the coverage check. However, the net error in total acres for the same group of farms was only 2.8 percent.

In addition to small farms, those most frequently missed in the census were farms in and around urban areas, farms with nonresident operators, and farms started during 1964, the census reference year. Errors in reporting total acres (if the coverage check figure is taken as correct) occurred most frequently on large farms, farms with several landlords and farms which acquired or gave up one or more tracts during 1964.

Comparison of coverage check results with both edited and unedited census values for total land and cropland harvested showed that net changes resulting from clerical and computer edit operations were quite small in relation to the estimated reporting bias for these two items.

F. Associated Census Evaluation Work

The coverage check was only one of two studies designed to evaluate the quality of information collected in the 1964 Census of Agriculture. A report on the other study called the "Panel Evaluation Survey," will be published as chapter 2 of volume 3, part 7. The basic difference between the two was in the types of items investigated. The panel survey was aimed at items for which it was believed that data of sufficient accuracy for evaluation purposes could only be collected by periodic contacts with respondents throughout the census reference year, instead of in a one-time survey characteristic of the coverage checks. The items included in the panel survey were sales, purchases, and inventories of cattle, hogs, and sheep; production and sale of corn, oats, and alfalfa; and selected farm expenditures. The 1964 Coverage Check was limited primarily to measuring the accuracy of census data on number of farms, land in farms, and cropland harvested.

G. Uses of Earlier Coverage Check Results

Starting in 1950, the results of the coverage checks have been made widely available to users of agriculture census data. This has been done primarily through publication in the regular census of agriculture volumes (appendix A, references 5, 6, 7, 8, and 9), but also through the preparation of special articles and reports (appendix A, references 3 and 4).

Preliminary results of the coverage checks have been made available to the Department of Agriculture as quickly as possible for use in revising current series on farm numbers, land in farms, cropland and major crops.

From the beginning, the levels of undercoverage shown by the coverage checks have been a matter of concern to the Census Bureau. Several procedural modifications have been introduced in the census in an attempt to improve coverage. Among these are--

1. In the 1945 and earlier censuses, the enumerator was given the farm definition and told to obtain questionnaires only for those places qualifying as farms. In 1950 and subsequently, the enumerator was instructed to obtain questionnaires for all places with specified types of agricultural operations. Decisions as to which of these places were farms were made during the processing of the questionnaires in the central office. This procedure was adopted in an effort to improve the coverage of marginal operations which, according to the 1945 Evaluation Survey, had accounted for a large proportion of the missed farms.

2. Another measure designed to improve coverage was the use, in the 1954, 1959, and 1964 censuses, of a listing book for each ED (enumeration district). The enumerator in each rural ED was instructed to record in his listing book the location and identification of every dwelling and every place with no dwelling but with agricultural operations, provided it was partly or entirely located in his ED. The enumerator was paid for each listing and therefore had an added incentive to canvass his ED thoroughly.

3. Starting in the 1954 census, a Township Sketch form was used in selected counties in an attempt to improve coverage, especially of nonresident operations which had been shown in the 1950 Evaluation Survey to account for a disproportionately 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 form, which was divided into squares, each equivalent to a quarter section, or 160 acres; and to identify each bounded area with the corresponding questionnaire or listing number. The procedure has been restricted to counties which are in that part of the country covered by the rectangular public land survey system and which are known to have relatively high proportions of nonresident operators. Cost and the difficulty of sketching the smaller and more irregular tracts which predominate in areas not covered by the survey system precluded the use of this procedure for the whole country; however, its use in areas covered by the survey system was expanded in the 1959 census and again in the 1964 census.

4. For all of the censuses covered by this discussion, enumerators have been provided with lists of large and special farms in their areas. These lists are assembled from many sources, but primarily from the previous census and the Department of Agriculture. The enumerator is required to obtain a questionnaire for each of these farms in his area or to explain why none was

needed. For the 1964 census, the number of farms included on these lists was substantially greater than in previous censuses.

5. Starting with the 1954 census, special procedures were adopted in selected counties in California, Florida, and Texas to facilitate a more complete and accurate enumeration of citrus groves.

6. In the 1954, 1959, and 1964 censuses, farm operators with fewer than 20 fruit and nut trees and grapevines were not asked to provide data on number of trees or vines and production for each variety. This was a direct result of findings from the 1950 Evaluation Survey, which showed that a large proportion of the operators with small numbers of trees and vines had failed to report them in the census.

H. Implications for Future Censuses

During the same period, however, there have been factors opposing coverage improvement. A few of these have been procedural, such as the decision not to use a special landlord-tenant questionnaire in the 1964 census. However, external influences may have had a more important effect. Among these have been the decreasing number, in many areas, of persons qualified and available for enumeration work, the gradually disappearing demarcations between suburban and rural-farm areas, and the increasing involvement of farm people in nonfarm employment and other off-farm activities, which makes it more difficult for enumerators to find them at home.

To the extent that coverage check results accurately reflect trends in coverage from 1950 to 1964, the net effect of these opposing factors on completeness of coverage appears to have been relatively small. Coverage may have deteriorated slightly over this period.

Coverage check results have shown that these continuing problems of undercoverage are largely concentrated among small farms which contribute little to total farm production. Another study[4] which made use of materials from the 1964 coverage check, has indicated that equal or better coverage of the economically significant farms could be achieved by mailing census questionnaires to lists of potential farm operators. In combination, the results of these studies have been a key factor in the decision to switch to a mailout/mailback approach for the 1969 census.

---

[4] A brief description of this study, called the IRS match project, is presented in section VIII, "Special Studies in Connection With the Coverage Check."

Section II. 1964 Census of Agriculture

As background to the coverage check, a brief description of the census itself is included here.[1]

In the United States, a census of agriculture is conducted every five years, in the years ending in 4 and 9. In 1964, the enumeration began on various dates between November 9 and November 23, except in Alaska, where it began on October 5. The beginning date for enumeration was set to follow the close of the harvest season in each State or portion of a State, and the enumeration was usually scheduled to be completed within 1 month from the date started.

A. Data Collection

The census is an enumeration of the agricultural operations carried on by individual persons, partnerships, corporations, and institutional units. The information collected provides data on the number of farms or agricultural operations in each county, each State, and for the Nation. It also provides an inventory of many of the basic agricultural resources and measures of agricultural production. Principal items investigated include agricultural lands, crop acreage and production, livestock numbers, livestock products, farm equipment, expenditures, receipts, and characteristics of farm operators.

Prior to the census enumeration, agricultural questionnaires were distributed by rural mail carriers to rural households in most areas of the country. A letter included at the beginning of the questionnaire requested the recipient, if he had agricultural operations of specified kinds, to fill in the questionnaire and hold it for the census enumerator, who would be calling for it in a few days. Approximately 8 million questionnaires were distributed in this manner; this was considerably more than the approximately 3 million farms in the United States.

A separate version of the questionnaire was used for each State, except in Texas, where two versions were used. The principal differences among the questionnaires for the various States were questions on crops harvested. Other differences included questions relating to fallow cropland, irrigation, forest products, and crops fertilized.

For enumeration purposes, the country was divided into 22,899 EA's (enumerator assignments),[2] each comprising an area which one enumerator could reasonably be expected to cover within 3 to 4 weeks.

In addition to the permanent staff members of the Washington and regional offices of the Census Bureau, the field organization of the census included 117 agriculture field assistants (supervisors), 1,800 crew leaders, and 23,000 enumerators. Basic training was given to enumerators and crew leaders in the form of a step-by-step home study program designed to present instruction in small parts and requiring approximately 14 to 16 hours to complete. In addition, crew leaders were given a 4-day training course which included actual enumeration, the completion and review of agriculture questionnaires and of other forms used by both enumerators and crew leaders, and the completion of a number of prescribed tests designed to measure the crew leaders' knowledge and understanding of the instructions, procedures, and assignments.

A listing form (see Form A2 of appendix B) was used to help the enumerator decide when to fill in a questionnaire. In rural ED's (where most farms are located) the enumerator was instructed to canvass his area completely, making a listing for--

1. Each occupied dwelling in the assigned area.
2. Each place on which there had been agricultural operations at any time in 1964.

With a few exceptions, each enumerator was provided with a map of his EA. As he recorded a dwelling or place on his listing form, he indicated its location by "spotting" it on his map, i.e., placing at the approximate location on the map, the number of the line on which the dwelling or place was listed on the listing form.

Enumerators were not given the precise definition of a farm, but instead used certain broad criteria to determine whether to obtain a questionnaire from the persons listed. Specifically, the enumerators were to fill a questionnaire for--

1. Every place considered by its operator to be a farm.
2. Every other place which had at least one of the following agricultural operations at any time during the year (except where the operator lived in another EA and also had agricultural operations at that location):
   a. Any cattle, 4 hogs or more, 30 chickens or more, or 30 turkeys or ducks or more were kept.
   b. Any grain, hay, tobacco, or other field crops were grown.
   c. A combined total of 20 fruit trees, grape vines, and planted nut trees or more were on the place.
   d. Any vegetables, berries, or nursery or greenhouse products were grown for sale.

In ED's located in urban areas and other areas with high ratios of dwelling units to farms, some modifications of the above procedures were made. A listing of all dwelling units was not required in built-up residential areas. For some ED's, operators of farms enumerated in the 1959 census were listed on the listing form prior to enumeration and the enumerator was required to visit and enumerate or otherwise account for each of these places. He was instructed to locate and

[1] For a more detailed description of the census see appendix A, references 8 and 9.

[2] Each EA (enumerator assignment) was made up of one or more ED's, the ED being the basic geographic unit for enumeration.

enumerate any other places with agricultural operations in his ED.

B. Processing the Data

Most of the processing of the data from the 1964 Census of Agriculture was done on electronic computers. However, some precomputer operations were necessary. The principal processing operations were:

1. Prepunch processing. This phase included the following steps:

a. An initial review of each enumerator's work, primarily to verify the amount of payment he was entitled to receive. This review included checks for completeness of all required forms and correct application of the various procedures. Some EA's were returned to the field for additional work.

b. An evaluation of county coverage to determine whether agricultural operations were adequately covered and to initiate correspondence to improve coverage where evidence indicated incomplete or unsatisfactory enumeration.

c. A precomputer editing and coding operation. The editing involved reviewing the questionnaires for selected types of errors, omissions and inconsistencies. Corrections were made in cases where the differences or discrepancies would otherwise have had a significant effect on published data. Some manual coding was required for items such as miscellaneous crops and kinds of livestock and poultry for which a specific question was not included on the questionnaire.

2. Card punching and card-to-tape operations.

3. Computer processing. The 1964 Census of Agriculture was the first U.S. agriculture census to make extensive use of large-scale computer processing equipment. The system for processing consisted of three main phases:

a. The first phase involved editing the raw data, imputing for certain specified classes of missing and inconsistent responses, and producing county totals for all data. This phase included the classification of farms according to size, tenure, economic class, type of operation, etc., and the calculation and application of weighting factors to produce estimates for those items which were collected only for a sample of farms. (See appendix B, reference 8, pp. XXVI-XXIX.)

b. The second phase involved making corrections arising from professional review of the results from the first phase, and producing a set of preliminary tabulations by counties. These were used for the preliminary reports which were published by counties, States, regions, and for the United States. The final product of this phase was a corrected, edited, weighted file of the detail data to be used in preparing final tabulations. Corrections which did not get included in the tape file were carried by hand as changes to the final tabulations.

c. The third phase consisted of preparing required tabulations for final publication.

Section III. Summary of Survey Design and New Procedures Used

This chapter describes the three samples used to obtain data for the coverage check, outlines the principal survey procedural steps, and describes two procedures used in the coverage check for the first time in 1964. A more detailed description of the survey design is presented in section VI.

A. Sample Design

The coverage check involved the use of two main samples plus a third sample which had a relatively minor role. The first, or *area sample*, was a sample of land segments which was selected to represent the entire area of the United States. Segments which were found to be located partly or wholly within urban places of 25,000 population and over were either eliminated from the interviewing completely or the part inside the city limits was eliminated if only a part of the segment was inside. The second, or *segment list sample*, was a sample of places enumerated in the 1964 Census of Agriculture. This sample also excluded places in urban areas of 25,000 population and over. The third, or *urban list sample*, was a sample of places in the urban areas not covered by the area and segment list samples.

To reduce costs, the area and list samples were designed to be overlapping so that, insofar as possible, the same farms would be included in both. The overlap involved having both samples associated with the same probability sample of land areas. The selection procedures for the sample of land areas are discussed in section VI. The nature and purpose of the three samples are summarized below.

The *area sample* consisted of the land areas included within 815 sample segments. An average segment contained about seven farms having all or part of this land inside the segment boundaries. This was a self-weighting sample with an overall sampling fraction of 3 in 4,000. The primary purpose of the area sample was to serve as a basis for estimating the number of farms missed in the census. However, it was also used to provide estimates of the number of farms correctly counted and underenumerated in the census. In addition, the area sample served to provide estimates for some types of overenumerated farms. (The terms correctly counted, overenumerated, underenumerated, and missed are described in section IV, "Concepts and Definitions.")

The *segment list sample* consisted of all places enumerated in the census whose locations had been marked within the boundaries of the 815 sample segments.[1] However, since there was considerable overlap between the two samples, only those list sample farms not contacted during the area sample canvass needed to be enumerated separately. The purpose of this sample was to obtain information for types of overenumerated places which would have been either highly difficult or impossible to detect using only the area sample.

The *urban list sample* consisted of a subset of persons reporting self-employment income from agriculture in the Bureau's Current Population Survey, plus a small selected subgroup not reporting self-employment farm income. The whole sample was restricted to urban areas having 25,000 population and over. The purpose of this sample was to cover those areas not covered by the area and segment list samples. This sample contained very few farms and its overall effect on the coverage check results was quite small.

B. Survey Procedures

In most surveys there is a single stage of enumeration followed by a processing operation. However, in the coverage check, several stages of data collection were interspersed with the preliminary processing stages so that the entire procedure became a more or less continuous process of obtaining and refining information reported on coverage check and census questionnaires.

The principal procedural steps involved in the enumeration and processing stages may be summarized as follows:

1. Initial (or Phase I) Field Enumeration. To obtain coverage check questionnaires for--
   a. All places having land in the area sample segments.
   b. All operators in the segment and urban list samples.
   c. Selected persons associated with places enumerated in a and b above.
2. Preliminary Processing.
   a. Central office review and edit of coverage check questionnaires.
   b. Searching and matching to identify associated census and coverage check questionnaires.
   c. Technical review, to classify area and list sample places as follows:
      (1) No followup required.
      (2) Mail or telephone followup required.
      (3) Field followup required.
3. Mail or Telephone Followup.
   a. To obtain additional data for coverage check farms for which no census questionnaire could be found.
   b. To determine the correct figure when the coverage check and the census differed on total acres or acres of cropland harvested.

[1] This marking operation, known as "spotting," is described in section II. In cases where the enumerator of an EA containing a sample segment failed to follow instructions with respect to marking his map, the "spotting" was performed by office personnel having no knowledge of segment boundary locations.

c. To obtain necessary additional information when this could be done with a few simple questions. (More complex cases were referred for field followup.)

4. Followup (or Phase II) Field Enumeration.
   a. To obtain additional coverage check questionnaires needed in connection with list sample cases.
   b. To resolve complex differences between coverage check and census results for a subsample of cases.
   c. To handle nonresponse to mail followup.

5. Final Processing.
   a. Editing and review of results of mail and field followup.
   b. Final search for census questionnaires matching coverage check questionnaires.
   c. Technical review to classify coverage check and census questionnaires by census enumeration status.
   d. Tabulation of results.

C. New Procedures: Weighted-Segment Approach to Estimation

In earlier coverage checks, only those area sample farms whose *headquarters* fell inside the sample segments were included in the final tabulations. The headquarters was defined as the residence of the operator, if he lived on the place, or the northwest corner of the place, if the operator did not live on the place. The *northwest corner* was defined as the furthest north of all points furthest west. If the headquarters of a farm were inside the segment, the entire farm was included in the coverage check estimate, regardless of how much land was outside the segment boundaries. On the other hand, if the headquarters were not inside the segment boundaries, the farm was not included in final tabulations, regardless of how much land was in the segment.

In 1964, all area sample farms having any land inside the sample segments were included in the coverage check estimate. These farms were weighted by the ratio

$$\frac{\text{land in farm in sample segment.}}{\text{land in farm}}$$

The principal reason for changing over from the headquarters rule used in 1959 to the weighted-segment approach in 1964 was that the latter method minimized the sampling variability of the missed farm component of net error. Concern over the sampling variability of this component of error stems from the fact that there does not appear to be any practical alternative to the use of an area sample for coverage of missed farms. Since farms of any size can be missed, it would have been possible, in some samples, for one or two extremely large missed farms to dominate the estimate. With the weighted-segment method it was less likely that this would happen.

The weighted-segment approach also served to control the sampling variability of the reporting error component of net error.[2] Previously, the variability of this estimate had been controlled by using a *large farm list sample* to represent all places having 5,000 acres and over. The weighted-segment approach eliminated the need for such a sample.

D. New Procedures: Subsampling Phase II Followup Cases

Although the number and average area of sample segments did not differ appreciably between 1959 and 1964, the weighted-segment approach increased the number of farms associated with each sample segment in 1964 by a factor of between 2 and 3. The number of cases requiring field reconciliation in 1964 was proportionately increased. Because of the large costs which would have otherwise been incurred, it was necessary to subsample selected groups of cases.

In general, there were three types of cases requiring field followup:

1. Census questionnaires in the segment list sample for which no coverage check questionnaire had been obtained.
2. Coverage check questionnaires for which no census questionnaire was found.
3. Coverage check and census questionnaires which represented the same agricultural operation but which had discrepancies greater than the prescribed tolerances.

All cases involving either a missing coverage check or census questionnaire were returned to the field for followup. These cases were very important since they represented farms which might have been either missed, underenumerated, or overenumerated in the census.

Thus, only those cases involving discrepancies between coverage check and census questionnaires were subject to subsampling. For this purpose, cases having differences which exceeded tolerance levels were divided into three groups, to be subsampled at rates of one (certainty, that is, no subsampling), one-half, and one-fourth. In general, higher subsampling rates were applied to those cases having the largest discrepancies.

---

[2] A description of the various components of net error is presented in section IV.

Section IV. Concepts and Definitions

A. Bias

The basic aim of the coverage checks has been to obtain measures of bias, as defined by

$$B = T - X$$

where T = desired or true value of a given statistic, such as number of farms or land in farms,
and X = expected or average value of that statistic over all conceivable repetitions of the particular census procedure used.

The bias component of error may be further broken down into *coverage* and *content* error, where coverage error results from the erroneous inclusion or omission of farms in the census, and content error is error in the data obtained for farms which were properly included in the census. Most of the estimates of bias presented in this report contain components from both sources of error.

B. Net Error

As a practical matter, there is usually no way to obtain unbiased estimates of the T's. Instead, estimates are made of the *net error*, defined by

$$Z = Y - X$$

where Y = expected or average value of a given statistic over all conceivable repetitions of the intensive collection procedures used in the coverage check.[1]

It was necessary, of course, to conduct the coverage check on a sample basis and to make estimates of the general form

$$Z' = Y' - X'.$$

Thus, the estimates of bias obtained from the study are themselves subject to the three kinds of measurement error:

1. Sampling variance, resulting from the collection of coverage check data on a sample basis.
2. Response variance, since both the census and the coverage check represented only one of all conceivable trials or repetitions.
3. Bias, since Z, the expected value of Z', is not equal to B.

C. Definitions Applicable to Coverage Check and Census Questionnaires

A coverage check questionnaire filled in as a result of the coverage check procedures was said to represent a *coverage check place*. If, in addition, the coverage check place qualified as a farm according to the census definition, it was termed a *coverage check farm*. A coverage check farm may or may not have had part or all of its land inside the segment boundaries. A *principal coverage check farm* was a place having at least some land in the segment.

Similarly, a census questionnaire (A1) filled in as a result of the regular census procedures was said to represent a *census place*, and if the place qualified as a farm it was termed a *census farm*. A census place in the segment list sample was called a *specified census place*.

The coverage check farm and its characteristics had the same definitions as used in the census, with one exception. The reference date for the coverage check determination of land in each farm was fixed at December 1, 1964, whereas in the census the reference date for each farm was the date on which the questionnaire for that farm was filled in. Differences between census and coverage check results occurring solely because of the difference in reference dates were probably quite small in relation to differences arising from other sources.[2]

The term "land in place" meant the total net acres reported either in question 7 of the census questionnaire or in section II, column 15 of the coverage check questionnaire.[3] If as a result of the clerical or computer edits a change was made in the census data, the edited census figure was used for comparison with the coverage check response. Also, if as a result of the reconciliation procedure a change was necessary in the coverage check schedule, the reconciled figure was used.

D. Control and Correspondence

The estimate of net error for number of farms does not provide any information about *gross* errors in coverage, e.g., the number of farms missed in the census or the number of places called farms in the census which were, in fact, not farms according to the census definition. For acreage items, an estimate of the net error by itself does not tell us how much of the error resulted from coverage check farms being missed in the census and how much was due to reporting errors for coverage check farms included in the census.

The concepts of *control* and *correspondence* were developed to overcome these deficiencies by providing a basis for associating individual coverage check farms and census farms. They were based on relationships between the land included in census farms and in coverage check farms. The definition of *control* was as follows:

1. A census farm whose land was not covered, even in part, by land in a coverage check farm, was

[1] A description of some of the more important coverage check procedures is presented in section V: Data Collection Techniques.

[2] See appendix A, reference 4, table 19.

[3] The coverage check questionnaire and relevant parts of the census schedule are reproduced in exhibits 3 and 1, respectively, of appendix B.

not controlled by any coverage check farm. Note: A census farm whose land was completely covered by land in nonagricultural coverage check places could not be controlled by any coverage check farm.

2. A census farm whose land was covered, at least in part, by land in a single coverage check farm, was controlled by that coverage check farm. Note: Census farms classified as "duplicates"[4] in the coverage check could not, by definition, be controlled by coverage check farms.

3. A census farm whose land was covered, at least in part, by land in more than one coverage check farm, was controlled by only one coverage check farm, determined as follows:

Rule 1. The coverage check farm covering the greatest amount of land.

Rule 2. If no one coverage check farm covered more land than all other coverage check farms, the controlling coverage check farm was determined by lot.

According to the preceding definition, a coverage check farm could control more than one census farm. The concept of *correspondence* was developed to designate, in such cases, one and only one of the controlled census farms as being associated with that coverage check farm. Specifically, the definition of correspondence was as follows:

1. If a coverage check farm controlled only one census farm, that census farm was designated as corresponding to the coverage check farm.

2. If a coverage check farm controlled more than one census farm, the one designated to correspond to the coverage check farm was determined according to the following rules:

Rule 1. The census farm whose acreage agreed most nearly with the acres in the coverage check farm.

Rule 2. If rule 1 did not provide a unique solution, the census farm that was in the same name as the operator listed for the coverage check farm.

Rule 3. If neither of the above rules provided a unique solution, the corresponding census farm was selected by lot.

E. Classification of Coverage Check and Census Farms by Census Enumeration Status

The concepts of control and correspondence were used to define a set of coverage classifications which in turn were used to analyze the positive and negative components of the *net error* for number of farms and other items.

A coverage check farm could be assigned to one of three classifications:

1. Correctly counted. Those coverage check farms controlling one or more census farms. Cases were separated into two categories:

a. Those coverage check farms controlling one and only one census farm.

b. Those coverage check farms controlling more than one census farm. Note: For such cases, one census farm was designated as the corresponding census farm; the remaining one(s), classified as overenumerated.

2. Underenumerated. Those coverage check farms covering part of the land in a census farm but not controlling any census farm. (This can occur when the census farm is controlled by a different coverage check farm.)

Two situations were distinguished:

a. Those coverage check farms for which the operator was listed in the census (but not enumerated).

b. Those coverage check farms for which the operator was not listed in the census.

3. Missed. Those coverage check farms not covering land on any census farm. Cases were classified into one of three subgroups:

a. Those coverage check farms not listed in the census.

b. Those coverage check farms which were listed in the census but not enumerated.

c. Those coverage check farms which were enumerated in the census but subsequently rejected as farms during the census edit.

A census farm could be assigned to one of two categories:

1. Correctly enumerated. Those census farms controlled by and corresponding to a coverage check farm.

2. Overenumerated. Census farms grouped under this classification were of two types:

a. Those controlled by but not corresponding to a coverage check farm. (See correctly counted farms, part b, above.)

b. Those not controlled by any coverage check farm. Three situations were distinguished:

(1.) The census farm was covered by a nonagricultural coverage check place.

(2.) The census farm could not be found in the coverage check.

(3.) The census farm was designated as a "duplicate."

Examples

1. A place enumerated in the census was incorrectly shown to have sufficient agricultural operations to qualify as a farm. This census "farm" was not controlled by any coverage check farm and was therefore classified as overenumerated.

2. A partnership operation was reported in the census on two separate questionnaires, one for each partner, and each accounting for half of the land. The coverage check questionnaire covering the partnership and the census farm designated to correspond to it were classified as correctly counted and correctly enumerated, respectively; the other census farm as overenumerated. The net error for land in farms was, of course, zero.

3. A farm operator received a census questionnaire in the mail, completed it, and gave it to the census enumerator who called for it. Two days later, another enumerator, who was out of his assigned area but did not realize it, obtained a second questionnaire from the operator's wife who did not know that a questionnaire had already been turned in. This duplication was not detected in the census processing. In accordance with the note given in

[4]For coverage check purposes, duplicate census questionnaires involved only those cases which were not detected during the regular census edit. In such instances, one census questionnaire was designed as the "original" census farm and the others were classified as "duplicates." The concepts of control and correspondence were applied only to the "original" census farm.

part 2 of the control definition, one of the two census farms was selected to be the "original" census farm and the other a "duplicate."

4. Mr. Adams owned 160 acres. In 1964 he had his own farming operations on 60 acres and the remaining 100 acres were rented out to Mr. Brown. Mr. Brown owned 20 acres in addition to this 100 acres. In the census, the questionnaire for Mr. Adams included all operations on the 160 acres he owned, and did not show any land being rented out. No census questionnaire was filled in for Mr. Brown. By application of rule 1 under part 3 of the control definition, Mr. Adams' 60-acre coverage check farm did not control his census farm, and was subsequently classified as underenumerated. Under the same rule, Mr. Brown's 120-acre coverage check farm controlled Mr. Adams' 160-acre census farm and these were subsequently classified as correctly counted and correctly enumerated, respectively. The resulting net error for farms and for land in farms was determined as follows:

a. Mr. Adams' 60-acre coverage check farm yielded

-1 farm and -60 acres

b. Mr. Brown's 120-acre coverage check farm yielded

0 farm and +40 acres

c. Net total: -1 farm and -20 acres

F. Components of Net Error

Using the classifications defined above, it is possible to rewrite the formula for the net error of a given statistic as follows:

$$Z = Y_M + Y_U + (Y_C - X_C) - X_O$$

where

Y_M = expected value of the statistic for coverage check farms missed in the census,

Y_U = expected value of the statistic for coverage check farms underenumerated in the census,

$Y_C - X_C$ = difference in the expected values of the statistic for correctly counted coverage check farms and for the corresponding correctly enumerated census farms,

X_O = expected value of the statistic for overenumerated census farms;

and

$Y = Y_M + Y_U + Y_C$ = expected value of the statistic for the coverage check procedure,

$X = X_C + X_O$ = expected value of the statistic for the census procedure.

Section V. Data Collection Techniques

A. Introduction

The main objective of the coverage checks has been to obtain measures of bias. This is done by using the best means available and feasible to collect accurate information from a sample of farms. These results are then compared with census results to determine the accuracy of the latter.

Despite the intensive procedures used in the coverage check, the results are not perfect. However, it is believed that the results are more accurate than those obtained in the census and provide a useful standard against which to measure the quality of the census enumeration.

Most of the data collection techniques used in the coverage check could theoretically be applied to the census itself if the resources were available. A few, however, are based on information or experience derived from the census. For example, preliminary coverage check results are compared with census results, and further investigation of differences is made when it is not clear why the difference occurred. Also, the best available census crew leaders and enumerators are hired as supervisors and enumerators in the coverage check and are given training more intensive than that provided for the census.

Some of the more important data collection and processing techniques are discussed in the following paragraphs.

B. Use of More Intensive Canvassing and Screening Procedures

As in the regular census, coverage check enumerators used a listing form to identify farm operators and agricultural operations. However, the listing operation in the coverage check differed from that used in the census in several important respects.

First, a more complete listing was made of persons associated with places having agricultural operations. Specifically, a listing was made for--

1. All household members 16 years of age and over for families living in the segment;
2. All household members 16 years of age and over for persons having control of land in the segment but not living there;
3. All persons associated with the agricultural operations of persons listed in 1 and 2 above such as landlords, tenants, and partners.

One reason for using a more complete listing is that farms missed in the census are frequently operated by persons who are associated with some other agricultural operation. For example, a farm operator's son living at home or a hired worker living in town might also have agricultural operations of his own which were overlooked during the census.

Second, the coverage check enumerator had to account for each tract of land inside his segment, including those tracts used for nonagricultural purposes, such as cemeteries, golf courses, school grounds, and airports.

Another difference between the coverage check listing form and the census listing form was in the screening questions used to determine whether the place had agricultural operations.[1] A comparison of the listing forms for the coverage check and census (appendix B, following detailed tables), shows that the screening questions used in the coverage check were more detailed and used lower cutoffs than those used in the census. For example, the question on income from sale of agricultural products used in the coverage check was not even asked in the census, and cutoffs were lowered from specified numbers of cattle and hogs in the census to any livestock in the coverage check.

C. Use of a More Detailed Questionnaire Limited to the Investigation of a Few Items.

The coverage check questionnaire was designed to provide detailed information on a few basic items, in the belief that such an approach yields more accurate data than that obtained by less intensive procedures. These data then, provide a standard against which the census observations may be compared. Thus, the questionnaire investigated the respondent's land, including its control and use, on a tract-by-tract basis, using a repetitive or probing type of questionnaire. In addition, information was obtained, if at all possible, from the person directly in charge of the operation even if this required additional callbacks.

To determine *net acres in place* the census questionnaire requested information on total acres for land owned, land rented or leased from others, land managed for others, and land rented or leased to others. (See appendix B, questions 3 to 7.) By contrast, the coverage check questionnaire investigated each *tract* of land in which the operator had an interest of any kind at any time during 1964. (See section II of the coverage check questionnaire, appendix B.) For each incoming tract, i.e., land owned by the operator, rented or leased by him from others, or worked on shares by him for others, etc., inquiry was made as to when he had acquired it and whether he still owned, leased, or worked it on shares. If the tract was not owned by the operator, the leasing arrangement was determined. Similarly, for each outgoing tract, information was obtained as to when this land was turned over to the tenant or sharecropper, when it was returned to the operator (if it had been), and what the arrangement was

[1] A coverage check questionnaire or a census questionnaire was filled in for each place having agricultural operations as determined by answers to the screening questions contained on the EPA2 or A2, respectively. This did not, of course, necessarily mean that the place would be counted as a farm.

or had been for the use of the land. The questionnaire was designed to determine from which of the incoming tracts each outgoing tract was taken.

This tract-by-tract investigation, coupled with the repetitive, probing nature of the questionnaire, was intended to reduce the likelihood of any land, either incoming or outgoing, being forgotten (or because of misunderstanding, being deliberately omitted) by the respondent. Thus the burden of deciding whether or not an individual was in control of and, therefore, should report on a given tract of land was shifted from the respondent to the interviewer and eventually to the office processing staff.

Although most of the questionnaire was devoted to obtaining data related to the land in the place, enough information on sales of farm products was obtained so that the characteristics of coverage check farms in terms of economic class and type of farm could be ascertained. Following the initial matching operation, if it appeared that a census questionnaire was never filled for the place, then one was obtained by mail or by a second interview in order to obtain the full range of census information about the place.

D. Use of Aerial Photographs and Sketch Maps

Aerial photographs were provided in most cases for use in determining exact segment boundaries and as an aid in preparing sketch maps of the land controlled by each respondent. In cases where the aerial photograph differed from the enumerator's county map with respect to boundary location, the aerial photograph boundary took precedence. Similarly, if a road, stream, or other segment boundary had changed since the aerial photograph was taken, the location of the old road, stream, or other segment boundary was used as the segment boundary. An example of an aerial photograph is presented in figure 1, which shows the boundaries of a typical segment and the segment designation number. In figure 2 the segment boundaries are shown on a county map corresponding to the same area.

Sketch maps of the land controlled by each respondent were prepared using a transparent overlay on the aerial photograph or a special sketch form in cases not covered by an aerial photograph or having tracts too small to be shown clearly on the overlay. As the enumerator determined the net acres in each tract (section II, column 15 of the coverage check questionnaire), he outlined the tract boundaries on his sketch map, showing it to the respondent, if possible, for the respondent's concurrence.

The purposes of the sketch map were to aid the enumerator in accounting for all land in the segment and to help estimate the acreages of tracts or parts of tracts. The latter purpose occurred most frequently in one of the following situations:

1. When the respondent did not know how many acres were in the tract;
2. When the respondent's answer did not seem consistent with the size of the sketched area;
3. When some of the acreage of a tract was both inside and outside the segment boundary, especially where the segment boundary was an imaginary line.

To estimate acreages, the enumerator used a transparent grid sheet having a scale equal to that of his sketch map. Depending on the scale, one small square represented a certain number of acres (usually 1) and each larger square some multiple of that acreage (usually 25). By counting the number of squares completely inside the boundaries, and estimating the acres represented by those squares partly in and partly outside the tract boundaries, an estimate of the total acreage could be obtained.

Section VI. Sample Design, Processing, and Estimation

A. Sample Design

As noted in section III, the area and segment list samples were both associated with the same sample of land areas. The design of this sample of land areas is described below.

1. Background considerations. Whereas the same sample of land areas was used in the 1954 and the 1959 evaluation surveys, by 1964 it was felt that this sample should be replaced. There were two principal reasons for this. First, the old sample was based on measures of size derived from the 1950 Census of Agriculture and, as a result, was likely to be less efficient than one based on more recent data. Second, it was felt that the burden of response associated with the coverage check procedures should be shifted to a new set of respondents.

When work started on the sample design it appeared desirable to select a sample which could serve, in addition to the coverage check and the panel evaluation survey described in section I, other major programs related to the 1964 Census of Agriculture. One of the more important potential uses of the sample was for a series of sample surveys on farm labor. Another possible use was for a sample survey to be taken 1 year following the census, as had been done following the 1959 census.

Consideration of the multiple uses of the new sample required taking into account the fact that a much smaller number of first-stage units (PSU's) would be needed for the evaluation surveys than for the sample surveys of agriculture and farm labor. It was therefore decided to select a relatively large sample of first-stage units from which a subsample would be selected for the evaluation studies.

Taking into account the sampling fraction used in the 1959 coverage check, the expected reduction in number of farms between 1959 and 1964, and various other factors, it was concluded that an overall sampling fraction of 3 in 4,000 would be appropriate for purposes of the coverage check. This sampling fraction constituted a 12-1/2 percent increase as compared with the overall fraction of 1 in 1,500 used in 1959. However, there was a compensating decrease due to the expected decline in number of farms from 1959 to 1964 which was estimated to be about 16 percent.

2. Sample selection. The sample of land areas used in the coverage check was selected in three principal stages as follows:

a. A set of 200 PSU's (primary sampling units) was selected, where each PSU was a county or group of counties.

b. Census EA's (enumerator assignments) were then selected within the 200 sample PSU's. Approximately 750 EA's were selected, each containing an average of about 140 farms.

c. Area segments were finally selected within EA's at a rate slightly greater than one per EA. The resulting sample contained 815 area segments distributed among 392 counties in 44 States.

3. Selection of primary sampling units. The results obtained from previous coverage checks suggested that the number of PSU's for the 1964 evaluation surveys should not differ greatly from the 196 which had been used in 1954 and 1959. However, to meet the other anticipated needs, the initial set of first-stage units selected consisted of 442 PSU's, from which a 200 PSU coverage check sample was drawn. To reduce the amount of work involved in the sample selection and also achieve certain other advantages, the PSU's were defined in much the same way as they were defined in the Bureau's Current Population Survey (CPS). That is, the approximately 3,100 counties and independent cities in the United States were grouped into about 1,850 PSU's, where each PSU consisted of either a single county or two or more contiguous counties.

Several kinds of controls were used in selecting the first-stage units. The major type of control involved a geographical stratification of units so that in general one PSU would be selected within each stratum. In addition, PSU selection was controlled in order to achieve a maximum amount of overlap with the Current Population Survey PSU's. This was done so that CPS interviewers could be used, if available, in the coverage check enumeration. The overall effect of this control was that 197 of the 442 PSU's selected were also in the CPS sample.

The sampling processes were also designed to control the balance between metropolitan and nonmetropolitan PSU's. Due to the greater difficulty of achieving complete coverage in congested areas, those PSU's containing Standard Metropolitan Statistical Areas were given probabilities of selection about 30 percent greater than would otherwise have been the case. On the other hand, probabilities for PSU's containing appreciable numbers of large farms (in terms of sales) were reduced because a separate one-stage sample of these large farms was to be selected for the agriculture and farm-labor sample surveys and because coverage of these economically significant places is somewhat easier to achieve.

The probability of selection for each PSU was based on a measure of size related to the extent of its agricultural activity. More specifically, after making an appropriate reduction for certain large farms (for the reasons indicated in the preceding paragraph) the assigned probabilities were made approximately proportional to the quantity $N_i \sqrt{V_i}$ where N_i was the 1959 number of

farms in the i-th PSU and $\bar{V}_i$ was the average value of farm products sold per farm in the i-th PSU (1959 census).

The distributions of the 442 and 200 PSU sample by census region are shown below.

Census region	PSU's in initial sample	PSU's in coverage check sample
Total.....	442	200
Northeast.....	38	14
North Central.....	202	82
South.....	136	80
West.....	66	24

4. Selection of second-stage units. At the second stage, EA's (enumerator assignments as defined for the 1959 Census of Agriculture) were selected within each of the 200 PSU's. It was planned that each EA selected at the second stage would have one segment selected within it at the third stage. However, if an EA was selected more than once at the second stage, an additional segment was selected within it at the third stage for each additional time the EA was selected. As will be explained below, the EA's were selected systematically; and as a result the only time an EA could be selected more than once was when its measure of size exceeded the sampling interval. In the great majority of cases, however, EA's were selected for the sample one time only.

The selection of EA's for the coverage check was accomplished by first taking a sample three times larger than that ultimately required. This process provided additional EA's for possible use in other surveys and also made it possible to select the coverage check subset from the following points of view:

a. There would be neither an over-representation nor an under-representation of EA's containing cities of various sizes.

b. The amount of travel required to go from area to area within the PSU would not be excessive.

EA's were selected systematically within PSU's according to the following steps:

Step 1. Determining a measure of size for each EA. Each EA was assigned a measure of size equal to the number of segments into which the EA would be divided if selected. The number of segments assigned to each EA was based on that EA's number of 1959 census farms. Using the census farm count for EA's, segments were defined to contain, on the average, 3-1/4 farms, except in the South, where each segment contained an average of 3-3/4 farms.

Step 2. Selection of EA's at three times the final rate. For reasons explained above, a sampling interval was calculated to yield a a sample three times that needed for the coverage check. The sampling interval within a PSU was made equal to the quantity

$$(1/3)\ (P_i)\ (4{,}000/3),$$

where P_i is the probability of selection of the i-th PSU. EA's were ordered according to EA number and their measures of size were cumulated. Then, starting with a random number between zero and the sampling interval, the first EA selected was the first one having a cumulative measure of size greater than the random start number. Successively selected EA's were those having the first cumulative measures of size greater than the random start number plus k times the sampling interval, where k was an integer equal to 1, 2, 3, ...

Step 3. Reordering of the selected EA's prior to subsampling. The selected EA's were reordered in such a way that the systematic selection of every third EA would prove satisfactory in terms of the points of view stated above, regardless of the start selected.

5. Selection of third-stage units. Within each EA chosen at the second stage, one or more segments (usually one) were chosen with equal probabilities. A two-stage procedure was used in order to minimize the amount of work required. The process involved an initial selection of "chunks" of land areas which were then divided into segments. This division or segmenting was generally accomplished with the aid of aerial photographs.

B. Searching and Matching

Following the phase I field work, a searching and matching operation was conducted. This procedure was carried out for each census of agriculture questionnaire in the segment list sample (called a specified A1[1]) and for each coverage check questionnaire containing land inside segment boundaries (called a principal EPA1[2]). The purpose of this operation was to match each specified A1 with an EPA1, although not necessarily a principal EPA1, and to match each principal EPA1 with an A1, again not necessarily a specified A1.

For the specified A1's the search was relatively easy since only the coverage check materials for the segment associated with the particular A1 were searched. For the principal EPA1's the search was much more involved. Here the census materials for the census enumeration area in which the segment was located, and for each adjoining enumeration area were searched. It should be noted that a census enumeration area was considerably larger than a coverage check segment and therefore took longer to search.

C. Final Search

Measures of the completeness and accuracy of census results depend heavily on how well the coverage check search operation succeeds in locating census questionnaires for which a complete or partial match did in fact exist. For this reason, a particularly intensive final search was conducted for the census questionnaires of all coverage check farms tentatively classified as missed.

The operation involved searching the census listing forms (A2's) for the name of the operator and any other

[1]The census questionnaire was popularly referred to by its form number.

[2]As with the census questionnaire, the coverage check questionnaire, was referred to by its form number.

person associated with him who had not been searched for during preliminary processing. If the coverage check place had 50 acres or more, the search included all persons connected with the person in charge, regardless of earlier search attempts. The census listing forms which were searched included those for the following types of EA's (enumerator assignments):

1. The EA containing the sample segment.
2. All contiguous EA's in the same county.
3. All contiguous EA's in other counties, any parts of which were within 5 miles of the sample segment.
4. Other EA's in which the person in charge lived or had other land.

For all places having 200 acres or more and for certain smaller places, an extended search was conducted if no A1 was found as a result of the above procedure. The extended search involved expanding the preceding search operation to include:

1. The remaining (unsearched) listing forms for the county containing the segment.
2. The census questionnaire folios for the entire county.
3. The listing forms and census questionnaire folios for any other county in which some of the land was located or in which the operator lived.

D. Grouping Related Coverage Check and Census Farms

To understand the need for grouping, consider the following two cases involving groups of related coverage check and census farms:

1. A single coverage check farm is enumerated in the census as two farms with the same total area as the coverage check farm, i.e., there is a net error (overcount) of one farm in the census, but no difference in total acres.
2. A place enumerated in the census as one farm is exactly equal to two coverage check farms, i.e., there is a net error (undercount) of one farm in the census, but no difference in total acres.

In case 1, the entire group--a coverage check farm and two census farms--would either be included or excluded from the sample estimate, depending on whether or not the coverage check farm was a principal coverage check farm, i.e., whether or not at least part of its land was inside the segment boundaries. If it were a principal coverage check farm, then both of the census farms it controlled would be included in the estimate--one as a corresponding census farm and the other as an over-enumerated census farm. There would be no contribution from this group to the estimated net error for total land, since the data for all of the farms in the group would receive identical weights.

In considering the case involving two coverage check farms and a single census farm, let

a = acres in the census farm;
A_1 = acres in the controlling coverage check farm;
A_{1S} = acres in segment of the controlling coverage check farm;
A_2 = acres in the noncontrolling coverage check farm; and
A_{2S} = acres in segment of the noncontrolling coverage check farm.

Then, the contribution of this group to the estimated net error for total land may be expressed as follows (the overall weighting factor of 4,000/3 can be ignored for this illustration):

$$\hat{d} = a\left(\frac{A_{1S}}{A_1}\right) - A_1\left(\frac{A_{1S}}{A_1}\right) - A_2\left(\frac{A_{2S}}{A_2}\right)$$

The problem for this group stems from the fact that although its expected value, i.e., its average value over all segments in the population, is equal to zero, its value for the individual segments containing these farms is generally a nonzero quantity. The zero expected value is due to the relationship

$$A_1 + A_2 = a$$

and the fact that the expected values of the weighting factors are the same; that is,

$$E\left(\frac{A_{1S}}{A_1}\right) = E\left(\frac{A_{2S}}{A_2}\right).$$

Thus, the group contributes to the variance of the estimate of net error in total acres even though it does not contribute to its expected value.

It should be clear that in the special case where A_{1S} is equal to A_1 and A_{2S} is equal to A_2, that

$$\hat{d} = a - A_1 - A_2 = 0$$

Thus, if both coverage check farms were completely inside the segment boundaries the estimated net error for total land would be zero and the need for grouping would not exist. On the other hand, if one of the coverage check farms was completely outside the segment boundaries, the estimated net error would be computed from

$$\hat{d} = -A_2\frac{A_{2S}}{A_2}$$

or

$$\hat{d} = (a - A_1)\frac{A_{1S}}{A_1}$$

depending, respectively, on whether the controlling or noncontrolling coverage check farm was outside the segment boundaries.

The types of cases discussed in the preceding paragraph resemble, in one respect, the 1959 coverage check situation where use of the headquarters rule[3] resulted in considering a farm as totally inside or outside the segment. In fact, actual grouping in 1964 was restricted to cases having one of the two coverage check farms in the group completely outside the segment. However, if the need had been recognized earlier, other types of groups would probably have been included also.

For purposes of obtaining a weighting factor, cases included in a particular coverage check group were combined into what was termed a "pseudo farm." Each

[3]See section III for a definition of the headquarters rule.

coverage check farm in the group was weighted by the quantity

$$\frac{\text{combined acreage inside the segment.}}{\text{combined total acreage}}$$

In this way the estimated (positive) net error in total acres for the census farm and coverage check farm controlling it, and the estimated (negative) net error in total acres for the underenumerated coverage check farm were equal, except for sign, and did not, therefore, contribute to the variance of the net error.

E. Estimation Procedure

Two basic types of estimates were made from the coverage check sample data. The first (referred to below as type I) was used for those items for which data were available from the full coverage check sample, such as total farms and number of farms by economic class, type, tenure, residence of operator and year began operation. The second (referred to as type II) was used for those items affected by the subsampling for purposes of the phase II followup to reconcile differences in total acres and acres of cropland harvested. Thus, type II estimates were used for total acres and acres of cropland harvested and for number of farms by size and by acres of cropland harvested.

1. Type I Estimates. With the sample design used, unbiased estimates of population totals could be obtained by adding together individually weighted cases from the area and segment list samples. Area sample cases were weighted by the quantity

$$\frac{4{,}000}{3} \times \frac{\text{land in farm in sample segment,}}{\text{land in farm}}$$

and list sample cases were weighted by the quantity

$$\frac{4{,}000.}{3}$$

It was then possible to combine these weighted cases to obtain, for any given item, an estimate of the net error,

$$Z' = Y_M' + Y_U' + (Y_C' - X_C') - X_O'.$$

Estimates of Y_M', Y_U' and Y_C', the coverage check totals for missed, underenumerated and correctly counted farms, were obtained from the area sample. The estimate of X_C', the census total for correctly enumerated farms, was based on census farms corresponding to coverage check farms in the area sample. The estimate of X_O', the census total for overenumerated farms, was based partly on overenumerated farms controlled by coverage check farms in the area sample and partly on overenumerated farms in the list sample, depending on the specific type of overenumeration involved.

Estimates of coverage check totals were obtained by adding the estimated differences, Z', to the published census totals, X, i.e.,

$$\hat{Y}_1 = X + Z' = X + \left[Y_M' + Y_U' + (Y_C' - X_C') - X_O'\right].$$

For total number of farms the component $(Y_C' - X_C')$ vanished since one correctly enumerated census farm, by definition, corresponded to each correctly counted coverage check farm. This was also true for farms classified by type and by economic class, since the census type and economic classifications were accepted for correctly counted farms. These classifications are determined on the basis of value of farm products sold, an item which was not investigated in detail in the coverage check. For farms classified by tenure, residence, and year began operation, however, it was possible for the census and coverage check classifications for corresponding farms to differ, so the factor $(Y_C' - X_C')$ was not, in general, zero.

Estimates of the component Y_M' for selected crop, livestock and poultry items, were computed using the data from census-type questionnaires completed for missed farms. For these items, estimates of the form

$$\hat{Y}_2 = X + Y_M'$$

are presented in section VII; however, it must be remembered that these estimates do not include components for underenumerated or overenumerated farms, nor for reporting errors on correctly counted farms.

2. Type II Estimates. The only cases subject to subsampling in the phase II followup operation were those having acreage discrepancies (total acres or cropland harvested or both) between coverage check and census questionnaires. Consequently, the only component of net error affected by the subsampling was that for correctly counted farms, i.e., $(Y_C' - X_C')$.

In order to make full use of the data available for correctly counted farms, a difference estimate was used for the items affected by subsampling. For a particular farm in the subsample there were, in general, three different responses (census, pre- and post-phase II) for the same item of information. Thus, for the j-th farm in the i-th stratum (where the stratum classification was based on the extent of difference between census and phase I results and determined the subsampling rates to be used), let

x_{ij} = the census result,

y_{ij} = the pre-phase II coverage check result,

z_{ij} = the post-phase II coverage check result,

and consider the differences

$$d_{ij} = z_{ij} - x_{ij}$$

and

$$e_{ij} = y_{ij} - x_{ij}.$$

Using these definitions, various estimates of census-coverage check differences may be constructed as follows:

$$d'' = \sum_i k_i \sum_j^{m_i} d_{ij}$$

$$e'' = \sum_i k_i \sum_j^{m_i} e_{ij}$$

$$e' = \sum_i \sum_j^{n_i} e_{ij}$$

where

m_i = number of cases in the i-th stratum included in the subsample,

n_i = number of cases in the i-th stratum subject to subsampling,

k_i = 1, 2, or 4 depending on the subsampling rate for the i-th stratum (i.e., included with certainty, one-half subsample, or one-fourth subsample).

By letting

$$d = E(d'')$$

and

$$e = E(e'') = E(e')$$

it is seen that d may be regarded as a more precise measure of the net error for correctly-counted farms, since it is based on the "final" findings of the phase II reconcilation following the phase I independent enumeration and matching, whereas e is based on the "preliminary" results from phase I and matching only.

Consequently, an estimate of the form

$$\hat{d} = d'' - (e'' - e')$$

was used to estimate the difference component $(Y_C - X_C)$ in th type II estimate. Here

$$E(\hat{d}) = d$$

and the sampling error of $\hat{d}$ was expected to be smaller than that of d'', because of the positive correlation between d and e.

Section VII. Coverage Check Results

A. Estimates of Net Error

Estimates of net error for number of farms, land in farms, and cropland harvested are presented in table 1, together with corresponding estimates from the 1954 and 1959 evaluation programs. Estimates of net error for number of farms by size, economic-class, acres of cropland harvested, tenure, and type of farm are given in tables 2 to 6, together with corresponding figures from the 1959 coverage check. The tables also include estimated totals, which are based on the sum of the estimated net error and the published census figure for each item.

The coverage check estimates of net error for most items include a component for reporting error on correctly counted farms. Thus, the estimates of net error for land in farms, cropland harvested, and for number of farms by size, acres of cropland harvested, and tenure reflect this component. However, estimates of net error for number of farms by economic class and type of farm reflect only errors resulting from over-enumerated, underenumerated, and missed farms.

The estimates presented in tables 1 to 6 are accompanied by estimates of sampling variability, expressed as standard errors. The chances are about 2 out of 3 that the difference between an estimate based on the coverage check sample and the figure that would have been obtained by applying the coverage check procedures to all farms would be less than the sampling error shown. The chances are about 99 out of 100 that this difference would be less than 2-1/2 times the sampling error.

The estimates of net error are also subject to nonsampling error, as explained in section IV, paragraph B. There are two aspects of the coverage check procedures which make it probable that the estimates of net error for number of farms exceed the true bias. First, the difficulty of carrying out searching and matching procedures is such that some of the census farms corresponding to coverage check farms may not have been located, especially in the case of smaller farms, for which less intensive search procedures were used (section VI, paragraph C). Second, once a census farm corresponding to a coverage check farm was located, there was no systematic attempt to search the census files further for duplicate questionnaires, so that some cases of duplication in the census may have been overlooked.

Some of the basic results which may be noted from tables 1 to 6 are as follows:

1. Relative net errors for acreage items are smaller than the relative net errors for number of farms. (table 1)
2. The estimated relative net error for number of farms was somewhat larger in 1964 than it was in 1959 and 1954. However, the estimates of relative net error for acres of land in farms were not significantly different for the 1964, 1959, and 1954 censuses. (table 1)
3. Relative net errors for number of farms by size were largest for "small" farms, whether one defines size in terms of land in farms, dollar value of sales, or acres of cropland harvested. (tables 2, 3, and 4)
4. For farms with sales of $2,500 and over (classes I to V), the estimated net errors for 1964 and 1959 did not differ significantly. However, for farms with sales of less then $2,500, the estimated net error was 19.3 percent in 1964 as compared with 13.7 percent in 1959.

It appears, therefore, that the increase in the estimated relative net error for total farms resulted primarily from less complete coverage of marginal farms in 1964. Farms in this category, i.e., those with sales of less than $2,500, accounted for only about 3 percent of the total value of farm products sold in 1964. The larger net error for marginal farms in 1964 may have resulted from one or more of the following factors:

a. The screening questions used in the 1964 enumerator canvass to identify persons with agricultural operations were less detailed and used higher cutoffs than those used in 1959. For example, horses, sheep, and goats were not mentioned in 1964; and cutoffs were raised from 20 to 30 for poultry and from 1 to 4 for hogs. Thus, farms with operations below the cutoffs used on the 1964 listing form but equaling or exceeding the 1959 cutoffs (e.g., a farm with 2 hogs or 25 chickens) had no chance of enumeration in the 1964 Census unless they answered "yes" to the question "Does this person... operate a farm (or ranch)?" In 1959 farms in the same category would have been enumerated on the basis of responses to the other screening questions.

b. The criteria used in the 1964 computer edit to determine which enumerated places should be retained as farms were more lenient than those used in the corresponding clerical operation in 1959. For example, a place with 10 acres or more was retained as a farm in 1964 if it had 50 chickens as compared with 100 required in 1959, or if it had 5 hogs and pigs as compared with 10 required in 1959. Places with sufficient operations to be retained in 1964 but not in 1959 were, in all probability, more likely to have been missed than other farms with sales of less than $2,500.

c. An increasing proportion of the marginal farms are located in areas which are primarily nonfarm residential. Farms in these areas are more difficult for census enumerators to locate.

B. Components of Net Error

Estimates of the components of net error are presented in tables 7 to 12. Table 7 shows these components by detailed "coverage status" classification for number of farms, land in farms, and average size of farm. By definition, there was one correctly enumerated census farm corresponding to each correctly counted coverage check farm. Consequently, for correctly counted farms, the census and coverage check estimates for number of farms are the same, while those for land in farms and average size of farm differ because of reporting error for these items.

In table 7, correctly counted farms controlling one census farm and those controlling two or more census farms are shown separately. Those controlling one census farm are further broken down according to whether or not there was agreement[1] on total land reported. Of the 3,026,000 coverage check farms controlling one census farm, 2,688,000 (88.8 percent) agreed on total acres. The corresponding figure for the 1959 coverage check was 89.1 percent.

Using the coverage check as a standard, an estimated 79,000 places were erroneously counted in the census. Of these 79,000 about 23,000 were not farms and the remaining 56,000 were parts of other farms. Of the 23,000 places which were not farms, 16,000 were places which did not have sufficient activity according to the coverage check to quality as a farm.

The coverage check procedure identified an estimated 456,000 farms which were missed in the census and an additional 23,000 which were underenumerated. Of the former group, 15,000 were enumerated in the census but rejected as farms in the census edit. For another 120,000, the operators were listed in the census but were reported as not having any activities requiring a farm questionnaire.

Tables 8 to 11 present estimates of the components of net error for number of farms by size, economic class, and acres of cropland harvested; and for total land, cropland harvested, and value of farm products sold.

The following results are among those considered most important:

1. Except for farms of 220 acres and over, the missed farm component accounts for most of the estimated total net error for farm counts. The missed farm component as a percent of the estimated total decreases as the size of farm increases (table 8).
2. For correctly counted farms, there was a tendency in the census to underreport the land in the farm as shown by the excess of census farms over coverage check farms in the first four size classes and the resulting excess of coverage check farms in the size class 220 acres and over (table 8). The overall net effect of this tendency may be seen in table 9, which shows a positive net error (census figure too low) of 2.8 percent for total land in correctly counted farms.
3. The positive net error for correctly counted farms of 3.5 percent for cropland harvested compared with only 2.8 percent for total land (table 9) is surprising because earlier coverage checks have shown cropland harvested to be more accurately and fully reported than total land. The corresponding estimates of net difference for correctly counted farms from the 1959 coverage check were 2.1 percent for total land in farms and 0.3 percent for cropland harvested. In the 1964 coverage check, data were obtained on a tract-by-tract basis. By contrast, a more intensive field-by-field procedure was used in 1959. It is possible that this shift in procedure resulted in an over statement in 1964 of total net error for cropland harvested. (See also comments in section VIII, p. .)
4. No attempt was made in the coverage check to measure reporting error for the individual data items used to calculate total value of farm products sold for correctly counted farms. However, data on sales were obtained from farms classified as overenumerated, underenumerated, and missed; and it is estimated that there was a relative net error of 2.9 percent for total value of farm products sold (table 9).

Census type questionnaires were completed in phase II of the coverage check for farms missed in the census. Estimates of the missed farm component of net error for crop, poultry, and livestock items are therefore available and are given in table 12, together with the census counts for corresponding items. While these estimates probably understate total net error, the missed farm components for these items are likely to contribute substantially more than other components to total net error. Therefore, it seems reasonable to indicate estimated minimum levels by adding the missed farm component to corresponding census figures.

For field crops other than tobacco and cotton, estimates based on the census total plus missed farm component are closer to corresponding U.S. Department of Agriculture (USDA) estimates than are census figures alone. With the further exception of corn, these sums are somewhat below USDA estimates.

The particularly close agreement of census and USDA figures for cotton and tobacco probably results from the special attention that was given these crops during the early stages of census processing. Records of acreage allotments for tobacco and ginning activity for cotton provided fairly accurate check data by county. Comparable county totals were obtained by clerks from the census questionnaires, and for some counties with large differences, lists of known producers were reviewed for possible omissions from the census. In addition, special efforts were made to eliminate card-punching errors that would have reduced the tabulated acres for these crops.

For livestock items the coverage check and USDA estimates are not directly comparable because of timing differences. The census inquiry related to inventories as of the date of enumeration which, for most farms, was

[1] Tolerances used to determine agreement on total land were as follows:

Acres in coverage check	Census and coverage check "agree" if difference does not exceed--
Less than 10	5 acres
10 to 99	50 percent of coverage check acres
100 and over	50 acres

sometime in November or December of 1964, the average date being December 2. USDA figures, on the other hand, are for estimated numbers on hand January 1, 1965. Differences between census and USDA figures may therefore reflect seasonal fluctuations in inventories arising from purchases, births, deaths, and movements of flocks and herds to other ranges, feedlots, and markets.

With one exception, the estimates for missed farms shown in table 12 varied from 1.6 to 16.7 percent of the census-total-plus-missed-farm component. The exception is chicken eggs sold, for which the missed farm component was only 0.5 percent of the combined total. Since production of eggs is concentrated on a fairly small number of farms with high value of sales, it might be reasonable to conclude that few, if any, of these farms were missed in the census. If this were true and the USDA figure were accepted as more accurate, then the difference would have to be attributed largely to reporting error for enumerated farms. On the other hand, if a few of the large producers were in fact missed, the estimated missed farm component would be subject to a relatively large sampling error; and an estimate similar to the one actually obtained would still be possible.

Table 7. Number of Farms, Land in Farms, and Average Size of Farm, by Coverage Status

Coverage status	Number of farms (thousands)		Land in farms (thousands of acres)		Average size of farm (acres)	
	According to census procedure	According to coverage check procedure	According to census procedure	According to coverage check procedure	According to census procedure	According to coverage check procedure
Total	3,158	3,559	1,110,186	1,181,706	351.5	332.0
Correctly counted[1]	3,079	3,079	1,096,741	1,130,189	356.2	367.1
Coverage check farm controlling one census farm	3,026	3,026	1,068,901	1,091,767	353.2	360.8
Total acres agree[2]	2,688	2,688	776,210	772,084	288.8	287.2
Total acres do not agree	338	338	292,691	319,683	865.9	945.8
Coverage check farm controlling more than one census farm[3]	53	53	27,840	38,421	525.3	724.9
Overnumerated (a census farm not corresponding to a coverage check farm)	79	-	13,445	-	170.2	-
Place was not a farm according to coverage check	23	-	1,416	-	61.6	-
Nonagricultural in coverage check	16	-	633	-	39.6	-
Not found in coverage check	5	-	89	-	[4]16.7	-
Duplicate	1	-	693	-	[4]520.0	-
Place part of another farm according to coverage check	56	-	12,030	-	214.8	-
Underenumerated and missed (coverage check farm not controlling any census farm)	-	480	-	51,517	-	107.3
Underenumerated	-	23	-	2,801	-	121.8
Operator listed in census	-	16	-	1,540	-	96.3
Operator not listed in census	-	7	-	1,261	-	180.1
Missed	-	456	-	48,716	-	106.8
Enumerated in census but rejected as a farm	-	15	-	924	-	61.6
Listed in census but not enumerated	-	120	-	12,260	-	102.2
Completely missed in census	-	321	-	35,532	-	110.7
Net error (total "according to coverage check procedure" minus total "according to census procedure")	-	401	-	71,520	-	-

[1]Estimates for correctly counted farms are based on weighted subsample (phase II) results adjusted to census levels.

[2]Total acres are considered in agreement if the census and coverage check figures do not differ by more than predetermined tolerance limits.

[3]When a coverage check farm controls two or more census farms, one of these is designated as the corresponding census farm. The total and average acres shown in columns designated "according to census procedure" are for corresponding census farms only.

[4]Figures are based on unrounded estimates for number of farms and land in farms.

C. Reporting Errors for Correctly Counted Farms

Tables 13 to 17 show, for various characteristics, estimates of correctly counted farms by class to which assigned in the census and in the coverage check. Elements on the main diagonal of each table are estimates of the number of correctly counted farms "identically" classified in the census and in the coverage check. The off-diagonal elements are estimates of farms classified differently, primarily as a result of errors made in reporting farm characteristics but also due to errors which occurred in recording and processing the data.

The reporting and other errors detected as a result of the coverage check may be broken down into components of response variance and bias. The response variance component represents the effect of errors which tend to cancel out where a large number of observations are made. Bias represents the effect of systematic errors which may occur, for example, as a result of the wording or format of a particular item on the census questionnaire.

The results presented in tables 13 to 17 may be used to compute measures or indexes of response variance and bias. These indexes are presented in table 18. The bias and variance measures are discussed in the following paragraphs.

1. Measures of bias. In estimating bias, the coverage check results are viewed as the standard against which the census results are compared. Their differences, then, are considered to be estimates of bias (net error) in the census. The census results are taken as an estimate of the average results which would have been obtained had it been possible to do independent repetitions of the census under the same general conditions. In like manner, the coverage check results are used to estimate the average results which would have been obtained if it were possible to do independent repetitions of the coverage check.

One purpose of tables 13 to 17 is to help clarify the nature of the net difference estimates presented in tables 8, 9, and 11. For example, table 8 shows an estimated net difference of -21,000 for number of farms in size class 10 to 49 acres. It may be seen from table 13 that this figure is equal to the difference between the coverage check total for farms in size class 10 to 49 acres (525,000) and the corresponding census total (546,000).

The bias or net difference relative to the estimated total number of correctly counted farms for which a response was obtained (or imputed) in both the census and the coverage check is called the net difference rate. This measure provides an estimate of the amount of bias in the census statistics. A positive value indicates a net deficiency in the census count; and a negative value that the classification was assigned to an excessive number of census farms. The algebraic sum of the net difference rates for a particular characteristic is zero. In the example for farms in size class 10 to 49 acres, the estimated bias is -0.71 percent (-21,000/2,963,000).

Another measure of bias is the index of net shift relative to the coverage check, which is obtained by dividing the net difference by the number in that class in the coverage check. For farms in size class 10 to 49 acres, this index has a value of -4 percent (-21,000/525,000), showing that the census estimate for this category was 4 percent greater than the coverage check estimate.

Net difference rates and indexes of net shift relative to the coverage check are given in columns (1) and (2) of table 18. These indexes show that there was a tendency in the census to underreport size for farms having 220 acres or more and those harvesting at least 50 acres of cropland. The indexes show substantial biases leading to undercounts of part owners and managers and of nonresident operators in the census, with corresponding overcounts in the other tenure and residence categories. Except for places starting operation before 1945, there may have been some tendency to update the year began operation as shown by the deficiency of census farms in classes prior to 1960-1964 and the resulting excess of farms in this latter class.

2. Indexes of response variance. One approach to the measurement of response variance involves the replication of some defined phase of the data collection or processing procedures and the subsequent comparison of results obtained for identical units. This method is useful for estimating the basic trial-to-trial variability in response, called simple response variance. Although the coverage check used an "improved" procedure, it provided data from a second source for identical farms, making it possible to compute estimates of simple response variance. Since the coverage check did not replicate the census procedure, however, resulting estimates of response variance are likely to be underestimates.

The results presented in tables 13 to 17 were used to compute indexes of response variability. The diagram below illustrates the approach for a particular size class.[2]

Census	Coverage check		
	In size class 10 to 49 acres	Not in size class 10 to 49 acres	Total
In size class 10 to 49 acres	a	b	a+b
Not in size class 10 to 49 acres	c	d	c+d
Total	a+c	b+d	n=a+b+c+d

The sum (b + c) relative to the estimated number of farms n in both the census and the coverage check, (b + c)/n, is called the *gross difference rate*, identified as g. The quantity g/2 is considered an estimate of the simple response variance. In order to make the estimates of response variability comparable from item to item, they are converted to an *index of inconsistency* identified as I. This is done by dividing the simple response variance, g/2, by the maximum value that it can assume, pq. The proportion of the population having the characteristic under consideration is p, and q = 1 - p.

[2]For a more detailed description see U.S. Bureau of the Census, *Evaluation and Research Programs of the U.S. Censuses of Population and Housing, 1960: Accuracy of Data on Population, Characteristics as Measured by Reinterviews*, Series ER60, No. 4. Washington, D.C., 1964.

For the first three characteristics presented in table 18, the average index of inconsistency is about 20 percent, and for the last two characteristics it is about 43 percent. These results imply that there is less error in reporting characteristics such as size of farm, acres of cropland harvested, and tenure than there is in reporting of characteristics such as residence or year began operation.

By way of comparison with results from the 1960 population census,[3] the average index of inconsistency for various age classes was about 5 percent and for number of children ever born, about 12 percent. Higher indexes were obtained for educational attainment classes and personal income classes, being 35 and 41 percent, respectively. The average index for residence in 1955 was 58 percent.

A final index of response variability presented in table 18 is the proportion of coverage check farms in a class which were identically reported in the census. With reference to the above diagram, this measure is simply the number of farms identically classified, divided by the number in the class in the coverage check, or $a/(a+c)$. When the proportion of farms in a particular coverage check class is small, this index can be used in the same way as the index of inconsistency.

Up to this point we have considered the effect of reporting errors on frequency distributions of correctly counted farms classified by various characteristics. Table 19 shows the effect of reporting errors on estimates of total acres in place and, for those cases where large differences occurred, indicates some reasons for the differences. The major question explored in this analysis is the extent to which large differences resulted from erroneous omission or inclusion of entire tracts, as compared with incorrect reporting of acres for correctly included tracts.

An estimated 56 percent of the cases with large differences in either direction involved erroneous inclusion or omission of entire tracts. Changes in control of the tract during the census year were apparently not a major factor in the erroneous omission of tracts. Of the 160,000 farms with large differences due to the omission of entire tracts only 24,000 involved a change in operator during 1964--for some or all of the omitted tracts. In about one-third of the remaining cases (46,000 out of 136,000) none of the omitted tracts was used for crops in 1964.

[3] Ibid., table 24.

Table 13. **Correctly Counted Farms by Size in Census and in Coverage Check**

(Thousands of farms. Data may not add to totals due to rounding)

Census size classification	Number of farms	Coverage check size classification											
		Under 10 acres	10 to 49 acres	50 to 69 acres	70 to 99 acres	100 to 139 acres	140 to 179 acres	180 to 219 acres	220 to 259 acres	260 to 499 acres	500 to 999 acres	1,000 to 1,999 acres	2,000 acres and over
Number of farms.....	2,973	149	525	195	285	336	332	206	171	441	193	89	51
Under 10 acres.....	161	132	24	2	-	3	1	-	-	-	-	-	-
10 to 49 acres.....	546	15	480	23	19	4	2	-	-	1	-	-	1
50 to 69 acres.....	191	-	13	149	12	11	1	1	2	2	-	-	-
70 to 99 acres.....	295	-	6	15	240	20	7	3	-	4	-	-	-
100 to 139 acres.....	329	1	-	6	14	273	23	5	2	5	-	-	-
140 to 179 acres.....	330	-	1	-	-	21	275	17	7	9	-	-	-
180 to 219 acres.....	217	-	1	-	-	3	17	169	10	14	3	-	-
220 to 259 acres.....	160	-	-	-	-	-	3	7	131	18	-	-	-
260 to 499 acres.....	429	-	-	-	-	-	4	5	18	378	22	2	-
500 to 999 acres.....	186	-	-	-	-	-	-	-	-	8	165	12	-
1,000 to 1,999 acres.....	83	-	-	-	-	-	-	-	-	2	2	74	5
2,000 acres and over.....	46	-	-	-	-	-	-	-	-	-	-	-	44

Table 19. Correctly Counted Farms by Type of Difference and Reason for Difference Between Census and Coverage Check Figures for Acres in Place

(Data may not add to totals due to rounding. Estimates are based on weighted subsample (phase II) results adjusted to census levels)

Type of difference and reason	Number of farms (thousands)	Land in farms (millions of acres)		
		According to census	According to coverage check	Net difference
Total	3,079	1,096.7	1,130.2	33.5
No difference in total acres	1,565	482.3	482.3	-
"Small" difference in total acres[1]	1,144	298.3	296.2	-2.1
Coverage check figure exceeds census figure	596	142.5	148.3	5.8
Census figure exceeds coverage check figure	549	155.7	147.9	-7.8
"Large" difference in total acres[1]	369	316.1	351.7	35.6
Coverage check figure exceeds census figure	257	152.2	220.2	68.0
One or more tracts omitted in census	160	83.0	133.9	50.9
Some or all of omitted tracts changed operator in 1964	24	11.5	16.2	4.7
Omitted tracts did not change operator during 1964	136	71.5	117.7	46.2
Omitted tracts not used for crops in 1964	46	25.5	47.6	22.1
Omitted tracts used, at least in part, for crops in 1964	90	45.9	70.1	24.2
Acres for included tracts incorrectly reported	97	69.2	86.3	17.1
Census figure exceeds coverage check figure	113	163.9	131.5	-32.4
One or more tracts erroneously included in census	47	47.9	31.9	-16.0
Acres for included tracts incorrectly reported	67	116.0	99.6	-16.4

[1]Differences are defined in terms of the size of the coverage check farm. A difference is defined as "large" if it exceeds:

5 acres for places having under 10 acres,
50 percent of the acreage for places having 10 to 99 acres,
and 50 acres for places having 100 acres or more.
A "small" difference is defined as one which is not large.

D. Other Results

1. Farm characteristics related to coverage and reporting error. Table 20 presents some results based on the coverage check area sample. Each farm in the area sample was classified in one of three "match status" classifications:

a. Missed in the census--No census questionnaire for a farm was found which accounted for any of the land in the area sample farm.

b. Enumerated in the census, complete match--A census questionnaire for a farm was found which accounted for essentially the same land as reported for the area sample farm, and the tenure classifications, i.e., full owner, part owner, manager, or tenant, were in agreement. Small differences between acres reported for the area sample farm and the census farm were allowed.

c. Enumerated in the census, partial match--All other cases for which some or all of the land in the area sample farm was accounted for on one or more census questionnaires for farms.

In table 20, the percentage distribution of farms by match status classification is shown for several different characteristics of farms and farm operators. These results show that several factors are associated with failure to find and enumerate farms and, once a farm is located, with failure to identify correctly the land included in the place.

Some types of farms most frequently missed were small farms (32.2 percent of the farms under 10 acres and 24.0 percent of the farms with 10 to 49 acres), farms in enumeration districts which were urban in character (25.3 percent), farms with nonresident operators (24.7 percent), and farms started in 1964 (22.6 percent). Farms operated by part owners and managers were less likely to be missed (5.9 percent) than those which were either tenant-operated (15.0 percent) or owner-operated (16.3 percent).

During the initial coverage check interview, each farm operator was asked whether a census questionnaire had been filled for his place. Of those who answered "no" to this question, an estimated 61.2 percent had actually been missed in the census.

Comparison of "complete matches" and "partial matches" shows clearly that the difficulty of determining total acres correctly increased in proportion to the number of separate tracts in the place, in proportion to the number of landlords, and in proportion to size of the place. Land changes during the census year (1964) also increased the likelihood that the census enumerator would get an incorrect figure for total acres. Farms operated by individuals were enumerated correctly more often than those operated by partnerships and other organizations, and fully owned farms were more often described correctly than partly owned or managed farms, with tenant farms occupying an intermediate position.

2. Evaluation of edit changes. In the computer edit of census schedules, information was checked

for completeness; and related items were examined to determine whether they were consistent with one another. When the information was complete and consistent, no changes were made; when it was complete but inconsistent, one or more changes were made; and when information was missing, it was either imputed, left blank, or related information was deleted, depending on the specific circumstances.

The coverage check list sample was used to estimate the effects of the census computer edit on two items--land in farms and cropland harvested. This was done by comparing the preedit and postedit census figures for those farms. The results are given in tables 21 and 22, and may be summarized as follows:

a. Net changes were not significant. Increases were about equal to decreases for both items whether viewed in terms of farms or acres.

b. The number of farms with changes and the amount of change (gross) were small for both items, being somewhat larger for cropland harvested than for total land.

It should be noted that some of the differences between preedit and postedit census figures for farms in the coverage check list sample may have been due to factors other than the computer edit. The preedit figures were transcribed from census schedules during various stages of clerical and professional review so that any changes arising from the manual review, after transcription had taken place, would be confounded with changes made in the computer edit. In addition, errors arising during the transcription and subsequent keypunch operations would also affect the differences shown in tables 21 and 22.

In a related study aimed at measuring the effect of the computer edit, 1964 census preedit and postedit data for farms in a sample of 30 counties were compared. In this study, the preedit figures were those resulting after all hand editing was completed. Also, a special check was made to insure removal of all keypunch errors so that the resulting differences would reflect only the effect of the computer edit on input identical to what appeared on the questionnaires after hand editing.

In tables 21 and 22 the results of the 30-county study are presented for comparison with those from the coverage check. The tables show that the computer edit in the 30-county study had a small effect on both land in farms and cropland harvested, but especially so for the first item. It is possible that the very low figures for land in farms reflect a considerable amount of hand editing for this item.

Table 23 presents results of a comparison of census preedit and postedit data with coverage check data for correctly counted farms. The table shows that net edit changes were quite small in relation to estimated biases.

3. Basis for determining whether an enumerated place qualified as a farm in census and coverage check. In tables 24 and 25, results are presented by "basis for qualifying as a farm in the coverage check" and by "basis for qualifying as a farm in the census," respectively. The need for the distinction arises from differences in the application of the farm definition in the census and the coverage check, and not from differences in the definition itself. Before considering the results of this section it may be useful, therefore, to summarize the definition of a farm and to indicate how this definition was applied in the census and the coverage check.

The 1964 census definition of a farm was based primarily on a combination of "acres in the place" and the estimated value of products sold. The word "place" was defined to include all land under the control or supervision of one person or partnership at the time of enumeration and on which agricultural operations were conducted at any time during 1964. Agricultural operations included the growing of crops; the raising of domestic animals and poultry, and the production of other agricultural products. Control may have been exercised through ownership or mangement or through a lease, rental, or cropping arrangement.

Places of less than 10 acres were counted as farms if the estimated sales of agricultural products for the year amounted to at least $250. Places of 10 acres or more were counted as farms if the estimated sales of agricultural products for the year amounted to at least $50. Places having less than the $250 or $50 minimum estimated sales were also counted as farms if they could normally be expected to produce agricultural products in sufficient quantity to meet the requirements of the definition. This additional qualification resulted in including in the census count some places engaged in farming operations for the first time in 1964 and places affected by crop failure or other unusual conditions.

Computer facilities were used in applying the definition of a farm to places enumerated in the census. The computer editing specifications for this involved two distinct steps:

a. First, an estimate of the total value of farm products sold was computed for each census questionnaire. For some agricultural products, including most livestock items, subtotals were obtained by simply adding the dollar values reported for the various items. For other agricultural products, including most field crops, dairy, and poultry items, the values of sales were estimated by using state average prices multiplied by quantities reported on the census schedule.

b. When a place failed to qualify as a farm on the basis of estimated sales, the computer editing specifications provided for counting it as a farm if it contained 10 acres or more and met any one of the following criteria:

(1.) 2 acres or more of crop failure.

(2.) 5 acres or more, in total, of cropland pasture and improved other pasture.

(3.) 10 acres or more, in total, of woodland pasture and other pasture.

(4.) 5 acres or more of land in summer fallow.

(5.) 50 chickens or more 4 months old or over on the farm.

(6.) 5 sheep and lambs or more on the farm.

(7.) 5 goats and kids or more on the farm (for selected States having this question).

(8.) 5 hogs and pigs or more on the farm.

(9.) 5 cattle or calves or more on the farm.
(10.) 2 cows or more on the farm.
(11.) 0.2 acres or more of tobacco harvested.
(12.) 100 pounds or more of tobacco harvested.
(13.) 0.5 acres or more, in total, of vegetables and berries harvested for sale and land in orchards, vineyards, and planted nut trees.
(14.) 3 acres or more of hay harvested, provided there were no cattle on the farm.
(15.) 2 acres or more of corn harvested, provided there were no cattle and calves or hogs and pigs on the farm.

Places having less than 10 acres and sales under $250 were counted as farms if they met any of the following criteria:

(1.) 5 acres or more of crop failure.
(2.) 100 chickens or more 4 months old and over on the farm.
(3.) 10 sheep and lambs or more on the farm.
(4.) 10 goats and kids or more on the farm (for selected States having this question).
(5.) 10 hogs and pigs or more on the farm.
(6.) 10 cattle and calves or more on the farm.
(7.) 4 cows or more on the farm.
(8.) 0.3 acres or more of tobacco harvested.
(9.) 500 pounds or more of tobacco harvested.
(10.) 2 acres or more, in total, of vegetables and berries harvested for sale and land in fruit orchards, vineyards and planted nut trees.
(11.) 5 acres or more of corn harvested provided that there were no cattle and calves or hogs and pigs on the farm.

Insofar as possible, an attempt was made to duplicate, in the coverage check, the census procedure for determining whether a place qualified as a farm. Sales figures were obtained in the coverage check for major livestock, poultry, dairy, forest, and nursery products. Coverage check places exceeding the $250 or $50 minimum sales levels were counted as farms on this basis. If the place did not have enough sales but had 10 acres or more of cropland harvested, it qualified as a farm. Finally, coverage check places not having enough sales or cropland harvested were counted as farms according to the individual item criteria just described, using data either from a matching census questionnaire or from a census-type questionnaire completed in phase II of the coverage check for farms missed in the census.

Table 24 shows the distribution of area sample farms by basis for qualifying as a farm in the coverage check. It shows that correctly counted and underenumerated farms were more likely to qualify on the basis of either sales or cropland harvested than were missed farms. Completely missed farms not qualifying on the basis of sales were more likely to have 10 acres or more of cropland harvested than were other missed farms not qualifying on the basis of sales.

In table 25, estimates of farms are presented by basis for qualifying as a farm in the census. For correctly counted and overenumerated farms, necessary data for applying the census definition were available from the census questionnaires. For underenumerated and missed farms this information was obtained from the census-type questionnaires which were completed in phase II of the coverage check.

As can be seen from table 25, class VII farms, sometimes referred to as "part-time" farms, accounted for over half of the farms qualifying on the basis of potential sales. About 10 percent of the overenumerated, underenumerated, and missed farms were classified on the basis of potential sales whereas the corresponding figure for correctly counted farms is about 5 percent.

4. Results of phase II reconciliation. Unweighted results of the phase II followup for correctly counted farms are presented in table 26. This phase of the coverage check involved field reconciliation of cases having discrepancies between census questionnaires and initial coverage check (or phase I) responses. After subsampling, cases which could not be reconciled by correspondence or telephone were returned to the field for phase II followup.

Results of the phase II followup are presented in terms of "phase I results verified," "census results verified," and "other." Strict criteria were used for classification in either of the first two categories, requiring exact agreement between phase II and the corresponding coverage check or census results for both total land and cropland harvested. The first column shows that responses from 522 of the 826 followup cases differed from those obtained in the census and phase I of the coverage check for at least one of the two characteristics land in place and cropland harvested. At first glance this result may seem to indicate considerable unreliability in reporting these acreage characteristics. It is important to note, however, that the results are based on a subpopulation of cases whose acreages may be more difficult to report; as may happen for example with farms having several tracts of land.

Table 26 shows results of the phase II followup by size of farm in phase II and by type of respondent in phase I. With reference to the farm size classification, phase II followups verified either the phase I or census results for about 50 percent of the farms having less than 220 acres but for only about 30 percent of the larger size farms. For the results presented by type of respondent, the table shows that phase I responses were verified somewhat more frequently than were census results for farms having "operator" respondent in phase I and somewhat less frequently for farms having "other" respondent in phase I.

Section VIII. Special Studies in Connection With the 1964 Coverage Check

A number of special studies were conducted in connection with the 1964 coverage check. Among these were the following:

A. EPA-IRS Match

This study was conducted to test the feasibility of doing the 1969 Census of Agriculture by mail, using a mailing list developed primarily from records of Federal income tax returns. Names of operators of farms in the 1964 coverage check sample were matched against tax returns for 1963 and 1964. In order to preserve the confidentiality of census information, all of the matching was done by Census Bureau employees. For each operator, it was determined whether individual income tax returns (Form 1040) had been filed in 1963 and 1964 and whether these returns included a Schedule F, Farm Schedule. Tax returns for partnerships and corporations were also examined where appropriate.

The results of this study were instrumental in the final decision to adopt the mailout/mailback approach for the 1969 Census. The results also showed that some supplementation of the basic IRS lists would be needed for the Southern States. A further check of the unmatched cases against records of the Agricultural Stabilization and Conservation Service showed that their lists would be suitable for this purpose.

B. Comparison of Alternative Methods for Resolving Discrepancies Between Coverage Check and Census Questionnaires

Obtaining information to resolve discrepancies between coverage check and census questionnaires became especially critical in the 1964 coverage check. Use of the headquarters rule in earlier coverage checks resulted in only about three farms per segment entering the final tabulations. The weighted-segment approach used in 1964 brought in an average of about eight farms per segment. Cases requiring reconciliation increased proportionately.

There were two principal methods of acquiring phase II information to resolve cases. When the information could be obtained from answers to a few simple questions, followup was generally made by correspondence although some contacts were made by telephone. The letters sent were of a standard form, explaining the purposes of the census and coverage check but not disclosing information which had previously been reported. At the bottom of the letter, space was provided for the question(s) to be asked and for the respondents' answer(s). For example, a typical letter might contain the following questions:

> How many acres are in the tract of land you rented from Mr. T. A. Jones?
>
> _____ acres
>
> Were you renting this tract on December 1, 1964? Yes_____ No _____

Because obtaining information by field followup is quite expensive, correspondence was used whenever it appeared that this method would be successful in obtaining the needed information. If, however, the problem of reconciliation were more complicated, if the letter were not returned, or if the answer proved unsatisfactory, a field followup was indicated.

With respect to the questions on land in place and cropland harvested, possible third answers, i.e., those differing from both the census and the phase I coverage check results, were handled as follows:

1. If the new answer were within 10 percent of the initial coverage check figure, then the initial coverage check figure was accepted.
2. If the new answer were within 10 percent of the census figure, the census figure was accepted.
3. If the new answer were not within 10 percent of either figure, then the new figure was accepted.
4. If circumstances of the case justified it, the technician resolving the case could exercise his own judgment as to which figure to accept.

With the increase in number of cases requiring reconciliation, it became desirable to study alternative methods of handling discrepancies. Two such approaches were investigated on an experimental basis at the Statistical Laboratory of Iowa State University, which was in charge of the phase II field work in a number of Midwest and Southern States.[1] One experiment dealt with information obtained by correspondence, and the other investigated the feasibility of reducing or even eliminating the phase II followup by having technicians resolve cases in the office which, under the existing procedures, were designated for field followup.

Experiment 1. Letter results and associated decision rules vs. field followup.

The letter experiment consisted of 96 cases containing discrepancies in total acreage, or in cropland harvested, or in both, which had been resolved by letter. That is, the replies had been considered satisfactory and thus each case was considered completed.

The cases involved in the letter experiment were sent to the field along with those selected for phase II field folowup. Attached to each letter experiment case was a summary, explaining the nature of the problem, what the respondent had previously reported, the figure that had been accepted, and the information desired. During the field interview, the "true" figure was determined as accurately as possible. Usually, it was one of the three previous figures. These final figures were accepted as the standard against which all previous results were compared.

[1] These experiments have been described in appendix A, reference 1.

After the field results had been returned they were analyzed by different methods. In addition, different acceptance rules were investigated. These were substantially the same as the 10 percent acceptance rule except that other percentage figures were used. These percentage figures were 0 percent, 15 percent, and 20 percent. (The only difference between using the 0 percent acceptance rule and the letter results was that the 0 percent acceptance rule represented the technicians' judgment in certain cases where the circumstances justified it.)

The principal conclusions which have been drawn from this experiment are as follows:

1. Correspondence is a valuable technique for reconciling differences in census and initial coverage check responses.
2. It appears that the technicians must exercise their own judgment sometimes in interpreting the answers received.
3. Any acceptance rule between 5 percent and 15 percent appears to work very well, particularly for the item land in place.
4. Original coverage check data on cropland (for cases involving discrepancies between census and coverage check results) were inferior to census data.

Experiment 2 Technician estimates using phase I data vs. field followup.

The purpose of this experiment was to see how well technicians could estimate values for land in place and cropland harvested, using only the information available prior to phase II fieldwork. Thus any notes written on the phase I coverage check questionnaire explaining the nature of the operation could be used by the technicians in making estimates. The enumerator sketch maps were sometimes helpful for this purpose. Also, if the operator had been the respondent for only one of these forms, this information helped in deciding which form contained the more reliable information. Many times, however, it was necessary to "guess" at a final figure. In general, such guesses were bounded by the census and phase I coverage check figures.

The technician experiment involved 157 cases having discrepancies in total land and 215 cases having discrepancies in cropland harvested. Cases were drawn from the subsample previously selected for phase II followup. As in the letter experiment, data from the followup field interviews were available as a standard against which the technician estimates could be compared. The relatively few cases involving very large differences were removed from the sample in order to reduce the experimental error and because the cost of field followup for these few cases would not be prohibitive.[2]

The results of this experiment showed that the technician estimates represented modest improvement over the coverage check figures prior to phase II followup; i.e., they were, on the average, closer to the final coverage check figures. Because of the relatively small additional cost of making such estimates (if made at the time of technical review), further investigation of this method may be desirable.

[2]Cases having very large differences were those with a difference exceeding 1,000 acres for land in place or 500 acres for cropland harvested.