Chapter 5. Processing the Data

GENERAL PROCEDURE

Processing the data for the approximately 4.1 million report forms mailed for the 1969 Census of Agriculture was an immense and complicated task. The skills of the subject-matter specialists in analyzing the tabulations in their respective specialties had to be blended successfully with the expertise of the systems analysts, programmers, and other computer specialists to process mass data. This combination of subject-matter and data-processing expertise was vital in bringing together data from census reports to produce accurate and useful statistics.

Census report forms were returned by the respondents to the Bureau's facilities at Jeffersonville, Ind. There, a clerical work force, varying from 100 to 450 employees, with the assistance of professional staff sent from Washington headquarters of the Bureau, accomplished most of the processing operations that were not done on the computer. In general, these operations included receipt and check-in of returns, routing of forms and other pieces of mail to the appropriate work group (such as the correspondence section, editing section, census file number search section, etc.), screening and review of report forms, and, after computer editing, review of computer changes and corrections.

The general procedure for processing the returns for the 1969 Census of Agriculture was as follows:

1. Sorting of returns and check-in
2. Screening the report forms
3. Data keying
4. Computer processing
5. Review of computer editing
6. Preliminary tabulation of data
7. Review of preliminary tabulations
8. Final tabulation of county, State, regional, and U.S. tables
9. Final table review and disclosure analysis

PRECOMPUTER PROCESSING

Check-In

The purpose of the check-in operation was to maintain control of the census returns that had been received and to determine which cases needed to receive letters or forms in each of the four followups. The census file number and a status code showing whether a return was in scope, out of scope, pending, etc., were keypunched to magnetic tape for each case received. These tapes were sorted by file number and matched by computer to the master mailing list. For each followup, mailing labels were prepared for the cases that did not appear on the check-in tapes. The check-in operation, therefore, was critical for the control of the entire census, as well as for the maintenance of good public relations by sending followups only to those people who had not returned completed report forms or given an indication that they soon would.

In preparation for check-in, all pieces of mail underwent a review and several sorting steps to insure that each was handled properly. Detailed instructions and charts for each step in the sorting process were provided so that a clerk could determine the action required for each piece of mail received. Figure 3 displays the initial mail sorting procedure, as an example.

Figure 3. Initial Mail Sort
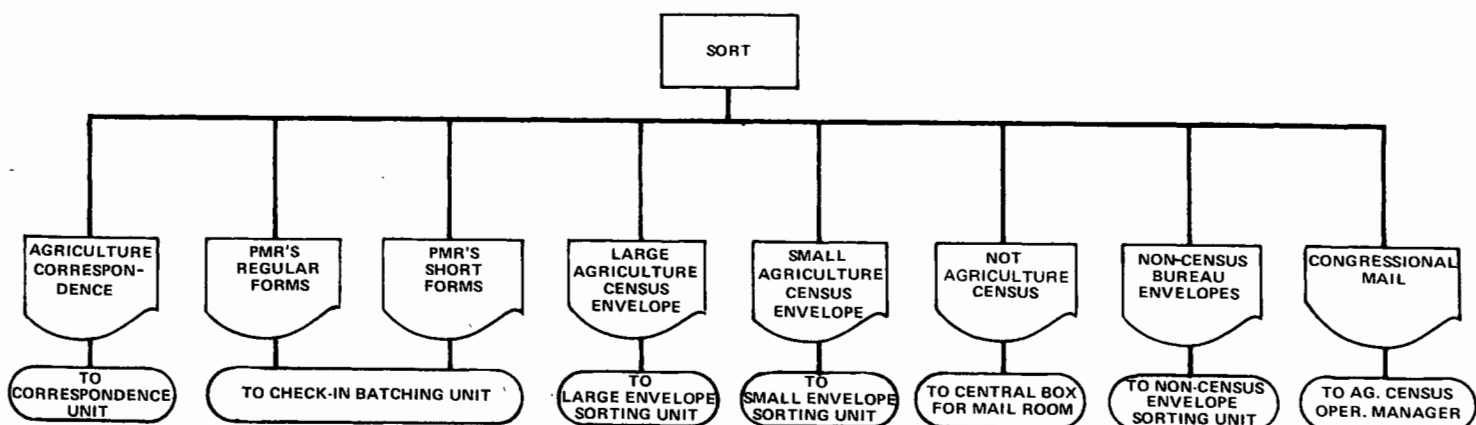
After the initial mail sorts, all mail was sorted into three main groups: Forms with correspondence, forms without correspondence, and correspondence only. The forms with correspondence were divided into two groups: (1) Those that did not require a reply (any significant comments were transcribed to the forms before they were sent to check-in); and (2) those that did require a reply. Report forms without correspondence were separated into three groups: (1) Those with no remarks in the section provided on the form; (2) those that were blank; and (3) those forms with remarks, which were separated into (a) forms not needing a reply, (b) out-of-scope forms, and (c) forms needing a reply. Correspondence without report forms was sorted into two groups. One group contained those letters without census file numbers. This group was checked clerically to a microfilm copy of the master mailing list to obtain census file numbers. The file numbers found were transcribed to the letters which were then checked in. Letters were sent to those respondents whose file numbers could not be located asking them to supply the Bureau with the file numbers. The other group of correspondence without report forms was further sorted into (a) requests for additional report forms, (b) letters from respondents who claimed that they had already filed a return, and (c) all other types of letters, such as requests for an extension of the time limit, requests for help in completing the form, questions about the census, etc.

One major problem which arose was a category of receipts referred to as multiple returns. These were instances in which the respondents returned two or more report forms, which were not yellow (indicating a multiunit), in one envelope. Many of these were cases of landlords and tenants returning their forms together, and families, such as a father and one son or more, returning every form they had received. All of these multiple returns had to be reviewed by a subject-matter specialist to determine which forms were in scope before they could be checked in.

Before the report forms and letters were sent to check-in, they were separated into categories: "Must cases" and QVS forms, out-of-scope forms, correspondence or blank forms, completed short forms, completed standard forms, PMR's (postmaster returns), and other. Once they were separated, they were batched into work units—100 standard forms (A1's), 200 short forms (A2's), 100 letters. These final sorting and grouping steps before check-in were taken so that the work could be sent directly from check-in keying to the appropriate work group.

Both check-in codes and followup codes were assigned to each work unit for check-in. The following check-in codes were used:

    0 Postmaster returns
    1 Correspondence only
    1 Blank forms
    1 Reports with correspondence without data
    2 Reports with correspondence with data
    3 In-scope reports
    4 Out-of-scope reports
    7 In-scope reports that had first been classified as out of scope and then determined to be in scope

This coding scheme was so constructed that a higher number superseded a lower number. Many cases, especially those that had been assigned check-in codes 0 or 1, were checked in more than once. These cases were considered to be incomplete, even

though in the case of those receiving code 1, something had been received from the respondent. The respondents were contacted and, when completed report forms were received by the Bureau, they were checked in again with a higher code. For example, if a respondent wrote to the Bureau requesting another report form, the census file number on his letter was checked in with a code of 1. His letter was then answered and when a completed form was received, the file number was again checked in with a code of 2, 3, 4, or 7, superseding the previous code.

Another set of codes was used to complete the check-in codes. This second set was used to indicate whether a followup action was to be taken, and if so, in which followup the case was to be included. Check-in codes 2 through 7 received a followup code of 0, which meant that no followup would be necessary since the cases with those category codes were not expected to be checked again. The followup codes used with category code 1 were as follows:

    2 Do not follow up this case until the second followup.
    3 Do not follow up this case until the third followup.
    4 Do not follow up this case until the fourth followup.
    5 Do not follow up this case until the fifth followup.
    7 Do not follow up this case (received too late for the last followup).
    8 Do not follow up this case (indefinite hold).

Once the work units had been prepared and the check-in codes assigned, the report forms and letters were sent to check-in data keying. Using electronic encoding equipment, the census file numbers and check-in codes were keyed to minitapes. The information on the minitapes was then pooled onto large "industry compatible" tapes. These "pooler" tapes were sent to Washington, D.C., where they were sorted and merged by census file number. In order to determine which cases were to be included in the followup mailings, the check-in tapes were compared to the master mailing list, and all census file numbers not appearing on the check-in tape as cases for which satisfactory responses had been received became eligible for followup.

After each check-in tape match a new master mailing list was generated. This new list carried the check-in codes for all addresses for which any type of response had been received, so that the latest status of each case was maintained. This allowed, for example, for a person from whom only a letter had been received to be excluded from the first followup in February, but included in the second followup if a completed report form had not been received by the beginning of April.

To evaluate the check-in process, a quality control check was made of a sample of one out of every 1,000 pieces of mail received, to determine prior to the cutoff date for each followup if any of these cases would be erroneously included in the followup.

Manual Edit

Once the report forms and correspondence had been checked in, they were routed to the appropriate work group. Explicit instructions were written for the different work groups and

given to each person within those sections. Whenever a problem arose which could not be resolved by the instructions, the individual referred it to his supervisor, and hence these cases became known as "referrals." The referrals occurred in every stage of the processing operations since it was not possible to make provisions for all the different types of problems which might arise. Frequently, additional instructions were written because the same type of problem kept arising. In other cases, the problems were resolved by subject-matter specialists. All report forms and letters from farm operations that were out of scope or suspected to be out of scope were sent to the subject-matter specialists for review and determination of their status.

The out-of-scope forms and letters were sent to the files and sorted according to the type of material—standard forms, short forms, correspondence—and filed in State and county code number order by type of material. No further action, other than conducting a quality control check on a sample basis to determine whether the cases were being reviewed correctly, was taken with the out-of-scope cases, although from time to time some of them were checked when they related to a problem with another case. The PMR's were also sorted by type of form—standard or short—and filed according to State and county code number. Some of these cases were included in a special remailing of PMR's conducted in late January 1970. (See chapter 4, page 38, for a discussion of PMR's).

All in-scope letters and forms with letters or remarks (except QVS and yellow "must case" report forms) were sent to the correspondence section for resolution. They were carefully reviewed, and form or tailored letters as needed were sent to the respondents as quickly as possible. (This process is discussed in detail in chapter 4, page 39.) When answers were received from the respondents, the appropriate action was taken (such as transcribing information to a report form, determining a case to be out of scope, etc.) and the forms and letters were sent back through the check-in and routing process to update the status of each case. Frequently, report forms from other sections of the processing operation were sent to the correspondence section for resolution of problems both prior to and after data keying.

The correspondence section was also responsible for searching for either a name and address associated with a census file number, or a census file number associated with a name and address. In many instances, either the mailing label came off the report form (see chapter 4, page 39), or the respondents either returned their file copies, which had no address labels or census file numbers on them, or wrote letters without supplying the census file numbers. The correspondence section used microfilm copies of the mailing lists (the original mailing list, the supplemental lists, and the lists of those included in each followup) to find the file numbers and the names and addresses. This was a particularly critical operation when searching for census file numbers since the cases could not be checked in without them, and therefore the respondents could not be kept out of the followups until a file number was found and some action could be taken.

The yellow ("must case") and QVS report forms were sent to the agriculture analysts for resolution. All "must cases" were briefly reviewed and held for shipment to Washington to be edited by the subject-matter specialists. The multiunit report forms were grouped according to multiunit and placed in folders until all reports for any one multiunit had been received. This provided for accurate control and followup procedures. The QVS forms were given to clerical personnel to transcribe all information reported onto regular A1 report forms. The transcribed A1 forms were stamped QVS, verified completely, and sent into the rest of the editing process. The QVS forms were shipped to Washington for review and analysis.

All other report forms were first sorted by State and then edited. This clerical edit and screening operation was done separately for the standard forms and the short forms.

The primary purpose of the clerical screening and editing operations was to determine (a) which report forms were ready for data keying to magnetic tape, (b) which had relatively minor problems that could be resolved in Jeffersonville by application of available editing procedures, correspondence, or telephone, and (c) which, if any, would have to be referred to Census Bureau headquarters in Washington. The procedures included screening out out-of-scope cases, identifying "must cases," and preparing the in-scope reports for data keying.

In the screening operation, clerks verified that all entries on the report forms were legible, made name and address changes for the files, made geographic coding changes, when appropriate, by correcting the numeric codes for the State or county, and performed some standard clerical operations. The work was divided into two parts. "General" editing consisted of a series of basic checks such as determining that the respondent's remarks or correspondence, if any, had been reviewed and resolved. "Specific" editing involved a detailed review of the minimum data required for a report to be considered complete. The editing was accomplished by clerks using various "editing guides" which provided specifications for both "general" and "specific" editing, such as which data items were "must" items (those which had to have entries before the report form was accepted for keying), what was considered minimum acceptable data, and when to refer report forms with problems that the clerks could not resolve.

The work of the editors was checked and verified, not only during the training period (a 10-percent random sample of a 200 report form work unit), but also after they had qualified as clerks (a 5-percent random sample of all work). In addition, the quality control procedure for the clerical editing provided for a 100-percent verification of the following, with the exception of referrals:

(1) All reports classified as out of scope were reviewed to insure that none were being eliminated which should be included in the census.

(2) All name and address change transcription sheets were verified to insure that these changes were properly made.

(3) All geographic coding changes were verified.

A quality control unit maintained individual records for each editor, and weekly summary reports were submitted to Census Bureau headquarters.

During the various editing and review operations, clerks encountered numerous problems which could best be resolved by

contacting the respondents. These problems (such as missing data, obviously erroneous or inconsistent entries, etc.) were usually limited to a few items which occurred frequently. In such cases, form letters could often be used. Therefore, the Bureau prepared a variety of form letters, and provided guidelines for selecting the appropriate letter to fit the situation. If no existing form letter was appropriate, or if the problem was so complicated that several letters would be required, the case was referred to a supervisor for a decision on whether it could be handled routinely, or would have to be referred to an analyst for resolution by "tailored" letter or telephone. Additional form letters were devised as needed.

All forms on buff paper with TOP codes 3, 4, or 5 (or 7, 8, and 9 as used in the supplemental mailings) which were determined by the editing or verification sections to be "must cases" were referred to the agriculture analysts. These were cases that (a) reported $100,000 or more in total sales or expenses, (b) reported enough crop or livestock production to amount to $100,000 in sales or expenses, (c) reported operating 50,000 acres or more, (d) appeared to be a report for an institutional organization because of a review of the address label or remarks, or (e) appeared to be part of a multiunit because of a review of the address label or remarks. After the review by the agriculture analysts, the forms determined to be "must cases" were stamped "must" and sent to Washington, along with all the other "must cases" which were on yellow report forms, for a comprehensive edit.

The edit performed by the subject-matter specialists in Washington on the "must cases" involved a complete review of all entries. It was considered extremely important that all information on these report forms be as complete, consistent, and error-free as possible since only a few incorrect entries or forms could seriously affect the data for any one county. These specialists checked and corrected, when necessary, the State and county codes in the census file number and made sure that all entries on the form were legible. Once all these editing

operations had been performed, the report forms were returned to the Jeffersonville facilities for data keying.

Data Keying to Magnetic Tape

After the report forms had been processed through screening and editing operations, they were repackaged into work units containing about 100 standard forms or 200 short forms and were forwarded to the data-keying section. The data were keyed directly onto magnetic tape by key encoder operators, one work unit to a reel.

Two different types of data were keyed—quantitative data (the entries provided by a respondent or, in some cases, an editor), and the applicable cell codes which were used to identify the items. The operators keyed only those items with entries and the applicable cell codes. The operators could ascertain the code for a particular data item by looking at the code numbers printed next to the item on the report form. This keying code generally consisted of a three-digit number designating the cell or field, but in some cases it was followed by subcodes, as illustrated in the examples below. Example 1 illustrates single-field data items, each of which had three-digit codes; example 2 illustrates multifield data items, which had a three-digit code for the first field and a one-digit subcode for each associated field.

The census file number was also keyed so that errors in processing could be traced and corrected for the final tabulations. The keying of the census file number was confirmed by the use of the check digit, the last number of the file number. When the entire census file number was keyed, the machine automatically computed a check digit. If the check digit keyed by the operator was not identical to the one computed by the machine, the machine "locked" to signal that there was an error.

EXAMPLE 1.

EXAMPLE 2.

The work of the keying operators was verified according to a quality control plan. This plan was divided into three stages on the basis of the operators' experience: (1) Training stage, (2) qualification stage, (3) process control.

Errors detected during verification were not corrected at that time, with the exception of those on "must cases." Instead, each work unit with an excessive error rate was rejected and rekeyed. All "must cases" were verified 100 percent and corrected.

If the number of report forms verified was—	Work unit rejected if number of errors greater than or equal to—
6	3
7	4
8-9	5
10-11	6
12-13	7
14-15	8
...	...
etc.	etc.

All rejected work units were rekeyed.

Training stage—All work units for each new encoder operator were verified 100 percent during the training stage while the operator was becoming familiar with the keying format. A keyer was allowed to advance to the qualification stage when he had keyed two work units with an error rate of 2.0 percent or less. Error rates were computed as the percentage of items in error, an item being either a data entry or a cell code on the report form. If the new operator failed on the first two work units he was allowed two additional work units. Keyers failing to advance after the second pair of work units were removed from the keying operation. All work units with an error rate greater than 2 percent were rekeyed.

Qualification stage—Keyers advancing to this stage attempted to qualify for process control. A 10-percent sample of reports from each work unit keyed was selected for verification. The decision to accept or reject each work unit was based on the number of report forms verified and the number of errors detected as shown in the table below.

If the number of report forms verified was—	Work unit rejected if number of errors greater than or equal to—
19-20	9
21-22	10
23-24	11
25-26	12
27-28	13
...	...
etc.	etc.

That is, for every two additional report forms in the quality control sample, one additional keying error was allowed.

A keyer was allowed to advance to process control when he had four successive accept decisions within a maximum of eight decisions. If he failed to qualify on the first sequence of eight work units he was allowed a second chance. Keyers failing to qualify for process control after the second sequence of eight work units were removed from the keying operation. All rejected work units were rekeyed.

Process control—After a keyer had advanced to process control, a 3-percent sample of report forms was selected for verification from each work unit keyed. As in the qualification stage, a decision was made to accept or reject each work unit based on the number of report forms verified and the number of errors detected as indicated in the table below.

A point system was used to determine if and when a keyer was to be removed from process control. The point system worked as follows:

1. Upon qualification for process control a keyer received two points as a bonus.

2. Each time a work unit was accepted, the keyer received one additional point.

3. Each time a work unit was rejected, he lost a point.

4. A keyer remained on process control as long as his point score did not reach zero. A reaccounting was made after each sequence of 10 decisions (a keyer with a point score of one or more began the next sequence of 10 decisions with two points).

5. If, at any time, a keyer's point score reached zero, he was placed back in the training stage and allowed to requalify under the same rules as before. If, after returning to process control, his point score again reached zero he was removed from the keying operation.

"Must cases" were keyed only by keyers who had qualified for process control.

These stringent control and verification procedures for the key encoding operations were instrumental in maintaining the overall error rate at less than 2 percent.

Each encoding machine held a small magnetic tape reel onto which the data were keyed. The small reels were then pooled onto a large reel which was transmitted via commercial airline to the Census Bureau headquarters in Washington, D.C. After keying operations had been completed and tapes transmitted, the report forms were held in the keying sequence pending receipt of confirmation from Bureau headquarters that the tapes were readable.

COMPUTER PROCESSING

Formatting

The first step in the computer processing of the 1969 Census of Agriculture report forms was to format the data into binary records. This computer record was a layout of 29 segments,

each of which consisted of one section or more from the two report forms.

The record layout contained both fixed-length segments and variable-length segments. A *variable-length segment* was made up of data for a series of items; an item was made up of data for a crop, for example, or for a type of equipment, etc. For each of the 15 variable-length segments there was an item code for each item designated within that segment. For example, segment 14 contained three items, and three item codes, as follows: 01, Irish potatoes; 02, sweetpotatoes; 03, tobacco. For any one item, the layout contained the item code and from two to six data fields; for example, for Irish potatoes the layout contained the item code and six data fields, as follows:

Word 1	Item code (01)
Word 2	Acres harvested
Word 3	Quantity harvested
Word 4	Acres irrigated
Word 5	Acres fertilized
Word 6	Tons of dry fertilizer used
Word 7	Tons of liquid fertilizer used

If one or more of these data fields (i.e., words 2 to 7) contained an entry, all of the data fields in the item were formatted (those fields that did not have entries were formatted as zeros), and if one or more of the items was formatted, the segment was formatted. The first word of each variable-length segment contained the segment number, the number of items in the segment, the number of words in the items, and the number of words in the segment.

Each of the 14 *fixed-length segments* contained only the segment identification and one set of from two to six data fields. If there was an entry for one or more of these data fields, the entire segment was formatted; those fields that did not have entries were formatted as zeros. The layout for fixed-length segment 26, for example, was as follows:

Word 1	Segment identification
Word 2	Total number of sheep and lambs on the place
Word 3	Total number of sheep and lambs sold
Word 4	Number of lambs under 1 year old
Word 5	Number of ewes 1 year old
Word 6	Rams and wethers 1 year old or older

This segment was present only in the records for farms having sheep and lambs.

The record segments were as follows:

Segment	Format	Description
1	Fixed	Identification
2	Fixed	Operator characteristics and recode
3	Fixed	Expenditure, other income, total market value of products sold
4	Variable	Machinery and equipment
5	Fixed	Land use
6	Fixed	Irrigation
7	Fixed	Artificial drainage
8	Fixed	Pastureland fertilized and conservation practices
9	Variable	Type of landlord
10	Fixed	Crops—short form
11	Variable	Corn and sorghums
12	Variable	Small grain
13	Variable	Beans, peas, peanuts, cotton, sugar beets, popcorn, broomcorn, and mint for oil
14	Variable	Irish potatoes, sweetpotatoes, and tobacco
15	Variable	Hay or grass
16	Variable	Field seeds
17	Variable	Other crops
18	Variable	Vegetables, sweet corn, or melons
19	Variable	Berries
20	Variable	Fruit trees, nut trees, and grapevines
21	Variable	Nursery or greenhouse products
22	Fixed	Livestock—short form
23	Fixed	Chickens, turkeys, and other poultry
24	Fixed	Cattle and calves
25	Fixed	Hogs and pigs
26	Fixed	Sheep and lambs
27	Variable	Other livestock
28	Variable	Contract or binding agreements
29	Fixed	Agriculture chemicals

The following operations were performed in the format run:

1. Crop production was converted into standard units of measure for those crops showing more than one unit on the form.

2. "Landlord only" and other types of out-of-scope records that could be identified in this run were separated from the data file.

3. Invalid codes were identified and classified, and appropriate action taken, as follows:

a. *Invalid State and county codes.* These records were printed out and dropped from the formatted file. The A1 or A2 report forms involved were corrected and then sent through the data-keying operation again.

b. *Invalid item codes.* These were codes not assigned anywhere on the report form. The identification, the invalid item code (cell code) and the associated data, and the next two item codes and their associated data, were printed out. Valid item codes that appeared out of sequence, including duplicates, were handled the same as invalid item codes. These item codes and all associated data were dropped from the record. Printouts of these records with bad codes were reviewed and the necessary corrections were made in a correction edit.

c. *Valid crop item codes that were invalid in a specific State or States* (for example, codes for cotton in Montana). These records were printed out and dropped from the data file. The A1 or A2 report forms involved were corrected and sent through the data-keying operation again.

Computer Editing

Computer editing is a mechanized process of screening, testing, and refining reported data; it essentially involves checking for reasonableness and internal consistency so that unusual information can be verified and corrected if necessary. In general, the computers are programmed to perform certain tests and make comparisons involving key ratios, such as acres harvested and yield. These key ratios were tested by comparing them against tolerance limits or parameters which had been derived from the previous census or current surveys. Computers were programmed to correct any item by a process of rounding, substituting a total by a sum of the detail or imputing on the basis of one of the several ratios in which the questionable component 'was contained, or by "flagging" the data entry for later inspection.

With hundreds of kinds of different crops as well as many different kinds of livestock operations covered by the 1969 Census of Agriculture, the computer edit programs were quite long and complex. The individual computer tests and checks amounted to several thousand steps, only a small fraction of which were required to edit the responses reported on any one form.

The mode selected for the communication of these computer edit specifications from the subject-matter specialists to the programmers during the 1969 Census of Agriculture was decision logic tables—tabular displays of all elements of a problem from conception to solution—reinforced by flowcharts and narratives. A total of some 888 pages of decision logic tables and directly related materials were prepared for the edit of the two forms (A1 and A2) for the general agriculture census. Several rounds of revisions were necessary in order to arrive at the desired precision and consistency.

Batch Edit

After the files were formatted, sorted, and merged by State, county, and census file number, they were divided by State, and the records were edited in two "batches" or more of all the records available at that time (instead of waiting until all of the records for a State or county were available before any computer editing was done for that State or county). For most States the computer editing was completed in two batches. The last batch edit was processed after the analysts' review of the "must cases."

The batch edits of the A1 and A2 forms included supplying missing entries, reconciling the acres reported for individual crops with the acres reported as total cropland, imputing production for crops when yield per acre was outside acceptable limits, and editing to assure consistency between and within the different sections of the report form. The edits also computed values for products sold, using average prices by State for each production item, and these estimated values were imputed if the reported value of products sold was outside acceptable limits.

In addition, the batch edit determined whether each record met the criteria for the standard A1 form, the short A2 form, or was out of scope, and classified the farms according to size, tenure, economic class, and type of farm.

The edit was divided into nine program groups, seven for the A1, one for both the A1 and the A2, and one for the A2. These program groups covered the editing of the sections of the report forms as follows:

Program Group 1 (A1):
a. Acreage and ownership plus crossline acreage
b. Land use
c. Reconciliation of crops and cropland harvested
d. Reconciliation of acres in orchard with detail of tree fruit acres

Program Group 2 (A1):
a. Irrigation
b. Drainage
c. Crops
d. Fruits and nuts
e. Vegetables
f. Nursery products
g. Fertilizer
h. Conservation practices
i. Forest products

Program Group 3 (A1):
a. Poultry
b. Cattle and calves
c. Hogs and pigs
d. Sheep and lambs
e. Other livestock

Program Group 4 (A1):
a. Estimation of TVP of livestock
b. Feed grain and hog requirement
c. Estimation of TVP of crops
d. Editing of reported TVP
e. Farm-related income

Program Group 5 (A1):
a. Operator characteristics and type of organization
b. Coding
c. Contracts

Program Group 6 (A1):
a. Machinery and equipment
b. Chemicals

Program Group 7 (A1):
a. Hired workers and expenditures
b. Value of land and buildings

Program Group 8 (A1 and A2):
a. Classification of record by type of report form (assignment of codes, weights, etc.)
b. Conversion of A1 to A2
c. Conversion of A2 to A1
d. Coding for A1 converted to A2
e. Failed edits and referrals

Program Group 0 (A2):
a. Acreage and ownership
b. Land use
c. Reconciliation of crop to cropland harvested
d. Cattle and calves
e. Hogs and pigs
f. Sheep and lambs
g. Poultry
h. Horses and ponies

Program Group 0 (A2)—Continued
   i.   Other livestock
   j.   Crops
   k.   Operator characteristics
   l.   Estimating and editing TVP
   m.   Farm related income
   n.   Coding
   o.   Machinery and equipment
   p.   Expenditures
   q.   Value of land and buildings

During batch edit, records that did not meet the minimum criteria for a farm were dropped from the data file and written on the out-of-scope file. The out-of-scope records from the batch edit were sorted and merged with the out-of-scope records from the format run. A listing was prepared of all out-of-scope census file numbers and sent to Jeffersonville for clerical review of the corresponding report forms to insure that they had been properly classified. A tabulation of the out-of-scope records by reason for being out of scope and by key items on the questionnaire was prepared for review.

A number of places that did not meet the definition of a farm (that is, they did not have 10 acres or more and at least $50 in sales, or, if they had less than 10 acres, did not have at least $250 in sales) were included in the census counts and were tabulated with the small farms (those with less than $2,500 in sales) because their inventories of livestock or their acreage and production of crops indicated that they would normally have had enough sales to be classified as farms. (All of them were included in the tables for either part-time or part-retirement farms.) For each record that did not meet the farm definition but was retained as a farm in the census tabulations, a criteria code was assigned to indicate the reason for including it in the census. If a record met more than one of the minimum criteria, the code for the first criterion satisfied was assigned to the record. (See appendix F, Table 1, "Farms Included in Census Count That Did Not Meet Farm Definition, by Reason for Including," and Table 2, "Farms Included in Census Count That Did Not Meet Farm Definition, by Number of Acres and Amount of Sales.")

In addition to determining which records were in scope, the computer edit also converted to A2 records those A1 records that were for farms that did not meet the criteria for A1's, and converted to A1 records those A2 records that did meet the criteria for A1's. Some A1 records were retained as A1's, and some A2 records were converted to A1's, even though they were for farms that had less than $2,500 in sales, because they met certain specified criteria and it was believed that they normally would have had sales in excess of $2,500. They would include new farm operations, farms having crop failure, and farms with large inventories and small 1969 sales. The Bureau decided, therefore, to include data for these farms in the tabulations for farms with sales of $2,500 (even though this affected the historical data comparability a little).

The specific criteria for retaining these records as A1's and for converting the A2 records to A1's, related to inventories, expenses, acres and production of crops, and cropland failure. Each of these records was assigned a criteria code indicating the reason that it was tabulated as an A1. The codes were assigned on the basis of the first criterion satisfied by the record. (See

appendix F, Table 3, "Farms With Sales of Less Than $2,500 Included in Tabulations of Farms with Sales of $2,500 or More, by Reason for Inclusion," and Table 4, "Farms With Sales of Less Than $2,500 Included in Tabulations of Farms With Sales of $2,500 or More, by Number of Acres and Amount of Sales.")

When information from A2 short report forms met the criteria for A1 report forms, the additional detailed information that would have been given on the longer form was imputed on the basis of responses for farms of similar size in the same area. Any A2 records on which more than $10,000 in sales was reported, and which therefore were converted to A1 records, were coded as "must cases." Records of the changes for these farms were printed out during the batch edit and sent to Jeffersonville for review.

The batch edit tabulated counts of the following key items for each county for records that (a) passed edit, (b) failed edit, and (c) were referred for review by analysts.

   1. Number of farms
   2. Total acres
   3. Cropland harvested
   4. Farms with 1,000 acres or more
   5. A1's
      a. Total
      b. "Must cases" identified prior to mail out
      c. "Must cases" identified prior to data keying
      d. "Must cases" identified during the batch edit
   6. A2's converted to A1's
      a. Total
      b. Acceptable—less than $10,000 TVP
      c. Must be reviewed—over $10,000 TVP
   7. A2's
   8. A1's converted to A2's
   9. Out-of-scope records
      a. Total
      b. Landlord, some agriculture
      c. Not a landlord, some agriculture
      d. Births dropped because they were also in the
         A2 sample

During the edit, records for A2's, which were sent to a 50-percent sample of small farms, were assigned weights of 2. The A1's, which were used for the 100-percent enumeration of all other farms, were assigned a weight of 1, except that those that had been converted from A2's and that indicated less than $10,000 in sales were assigned a weight of 2.

Batch Edit Correction

As a result of the batch edit, a "Failed Edit and Must Case Listing," printed on a high-speed printer, was produced for each "must case" and for each farm that had one item or more fail the edit program. These listings were referred to as farm records, and for each farm, the record displayed only the items that had failed the edit and the items which did not fail but were different before and after the edit. Each page of the listing contained the items for one farm, although some farm records extended to two pages.

The listings were separated, placed in State folios in lots of 500 consecutively numbered farm records, and then shipped to

Jeffersonville. The first step in the Jeffersonville operation was to match the listing to the report form file and pull the appropriate report forms. The listing sheets and report forms in Jeffersonville for the "must cases" were packaged and shipped to Washington, D.C., for review. The rest of the listing sheets and the corresponding report forms were sent to a technical analyst in Jeffersonville for review.

The purpose of the review of the batch edit was to process each farm record so that information going back through the computer would allow the computer to accept a changed "failed-edit" item, to add items, change items, delete items, or to delete the entire record for a farm. Subject-matter specialists in Washington, working on the "must cases," and technical analysts in Jeffersonville, working on all other cases, were responsible for performing this review.

In Jeffersonville, the technical analysts were trained by agriculture analysts. The agriculture analysts, who were responsible for the quality of the operation, assessed and insured the quality by inspecting work as it was performed and assisted on difficult cases. One clerk assisted each technical analyst by supplying the listing sheets and appropriate report forms, readying each listing for data keying, and refiling the report forms as work was completed.

The technical analyst reviewed each listing sheet and made appropriate decisions, that is, decided whether to change an item, delete the entire record, accept the changes made, etc. He indicated on the listing sheet the action to be taken by the computer with two sets of codes—general purpose codes and correction action codes. The general purpose codes indicated what was to be done to the entire farm record, while the action code indicated what was to be done to a specified item. These codes were as follows:

General purpose codes—
    0 Delete the entire record
    1 Make the indicated corrections and reedit the record
    2 Make the indicated corrections, but do not reedit the record
    3 No corrections, change a "failed edit" record to a "passed edit" record (so that it would not fail the edit program again)
    4 Make the indicated corrections, change a "failed edit" record to a "passed" edit record, and reedit the record
    5 Reedit the record and supersede code 2.
Correction action codes—
    6 Delete a record
    7 Add or replace a record
    8 Delete a segment or item
    9 Add or replace a data field

An item locator code was assigned to every location within the farm data record. These item locator codes were used in inserting corrections in the farm data file.

To "replace" a record meant to rekey the entire report form, and the new record generated took the place of what was already in the data file. If no corrections were needed for a record, a general purpose code of 3 was assigned. When there were a large number of corrections for a farm (25 or more), the form was corrected, rekeyed, and sent through the computer processing again. After the technical analyst marked the

corrections on the listing sheets, the clerk prepared the listings for data keying by underlining data that needed to be keyed and inserting missing locator codes. The corrections were keyed to tape, verified 100 percent and, when correct, were shipped to Washington, D.C., for matching to the data file on the computer. The corrected files were displayed and reviewed to ensure that the corrections had been made properly and to determine if further corrections were necessary.

Final Data Merge

After the batch edit, the corrected files for each State were merged into one file in sequence by State, county, and identification number. The following operations were also performed in this merge:

1. Unduplication. If there were two records or more with the same census file number, the first one was kept on the data file and the rest dropped. All the duplicate records involved were displayed.

2. Imputation counts. Farms by class and TOP code were tallied and used in the imputation of nonrespondents. (Imputation is discussed in the following section.)

3. Bad data records were displayed and dropped from the data file. Among these records were those with negative income and A2's with large acreage. These records were reviewed to determine if they should be included in the tabulations.

Imputation of Nonrespondents

After the files were corrected, merged, and unduplicated the next step was to impute data for the nonrespondents. The procedure used was, in effect, by duplication of the responses on a report form for another farm in the same county. This meant that all the data for the farm selected was given a weight of two, i.e., they were counted twice. The general rule for imputation for nonrespondents was:

Number of farms to be imputed (weight doubled) =
$$I = D\left(\frac{F}{M-P-D}\right) \text{where:}$$

    D Number of nonrespondents on the county mailing list at final closeout
    F Number of in-scope farms for the county before the summary diary stage
    M Number of addresses on the mailing list for the county
    P Number of postmaster returns received for the county

Farms for duplication (doubling of weight) were taken only from those records with TOP code 4 or 5 because a 100-percent followup had been done for cases with TOP codes 0,1,2, or 3. The records with TOP code 4 were selected from economic class group 2 to 5, proportionate to the distribution of the number of farms in these economic classes. These were farms with total value of products sold of $20,000 to $39,999; $10,000 to $19,999; $5,000 to $9,999; and $2,500 to $4,999, respectively. The records with TOP code 5 were selected from economic class group 6 to 8, proportionate to the distribution of the number of

farms in these economic classes. These were farms with total value of products sold of $50 to $2,499 and farms classified as other low income (part time and part retirement). The weights of the selected records in the data file were doubled. This meant that each record was counted four times, since a weight of two had already been assigned because the farm was in the 50-percent sample.

TABULATING THE DATA

General

After the batch edit and edit corrections had been completed, the data were ready to be tabulated. Two different types of tabulations and corrections were run first on the computer—diary tabulations (by county) and county tabulations for publication. The tabulations for publication were reviewed and, when accepted, were prepared for publication (see chapter 7). The diary tabulations provided data in a detailed format which were used as a resource for locating problems in the data for almost 3 million report forms. After both sets of tabulations were reviewed and farm records corrected and retabulated and the county data published, data for the State parts and the U.S. Summary were tabulated.

Diary Tabulations

The diary tabulations were divided into two tally programs. One tallied all the crop items from the A1 report forms at the county level. The other tallied all the noncrop items from both the A1 and the A2 report forms, plus the crop items from the A2, at the county level. Data were displayed for each item according to economic class, along with related data. These tabulations were reproduced on microfilm and used along with the change index for review of the county tabulations.

County Tabulations

Data tabulated by the computer had to be translated from magnetic tape to printed documents, or paper copy. The paper copy was produced by high-speed printers which were auxiliaries (or "output units") of the Census Bureau's computer systems. The tables that came from the high-speed printers as computer printouts were carefully reviewed by subject-matter specialists. This review consisted of three main steps: (1) Review of available comparative data for 1969 key items; (2) review of State and county tables by Statistical Reporting Service representatives from the State offices; and (3) review of county tables by Census Bureau staff.

Several tools were provided to the reviewer so that he might determine whether or not totals for a particular item were questionable. Within the tables being reviewed there was a substantial amount of intrarelated check data as well as counts from the 1964 Census of Agriculture for almost all items. In addition, U.S. Department of Agriculture estimates for principal items were utilized.

Because of its need for and use of data from the census of agriculture, the Department of Agriculture cooperated in the review of county tabulations by sending State representatives to

assist in this phase of the processing. Most States were represented by at least one SRS statistician, although there was only one for the New England States and one for Maryland and Delaware. The Texas office sent two statisticians because of the unusually large number of counties in that State. The SRS offices in Hawaii and Alaska were not asked to send anyone—instead, tables and written instructions for review were mailed to them for examination and returned to the Bureau of the Census with their criticisms. SRS statisticians commonly spent 1 or 2 weeks reviewing data for their respective States. On arrival at the Census Bureau, the SRS statisticians were given a brief orientation about the census covering (a) census procedures, (b) county table format, (c) diary table and change index format, and (d) their review responsibilities.

The separate reports to be published for each county demanded most of the SRS reviewer's time, although this was limited primarily to a review of the crop and livestock items. They were asked to provide marginal notes relating to any entry which they considered questionable on the basis of comparison with check data or personal knowledge, and to write their criticism on forms specifically designed for that purpose. Generally, SRS reviewers only identified potential problems. Initially, it was intended that SRS statisticians would use the diary tabulations as an aid in their review, and heading overlays were prepared for that purpose. However, the diary proved too complex to be used effectively by people coming in without preparation and for only a brief period.

After the SRS State representatives had finished their review, the census subject-matter staff reviewed all of the data and wrote specific directions for handling each item to be changed. The action taken by the Bureau staff included the following:

1. Reviewing the criticisms prepared by the SRS State representatives.

2. Independently reviewing to identify inconsistencies and potential errors, particularly for the items not reviewed by SRS statisticians.

3. Checking lists of large farms from current lists and from the 1964 Census of Agriculture to insure that these operations were included in county and State totals.

4. Obtaining reports from farm operators for large places that had not been included on the tabulations.

5. Identifying and correcting data-keying, reporting, and processing errors.

6. Identifying and eliminating duplicate reports.

7. Assigning correct State and county code numbers for large operations to insure that these operations were tabulated in the proper State and county.

The staff in Washington worked with individual report forms, which had been returned to Washington after the batch edit corrections, for the "must cases" and the very large farm operations, as well as with the county tabulations. Once the county table review was completed and the selected individual report forms checked, the criticisms were sent to Jeffersonville for a review of the problems involving the other report forms.

County Table Corrections

When the review of the county tables was complete, corrections were carried to the individual farm records in the same manner as they had been after the batch edit. After the corrections were made, the county tables were tabulated again for another review.

This second review consisted primarily of checking to see that the problems found previously had been resolved. If it was found that any corrections had not been made, either the tabulations were corrected on the computer or hand corrections were made directly on the printouts of the tabulations. The data file was corrected as often as necessary to assure the accuracy of the data.

Tabulations for States, Regions, Divisions, and the United States

Some of the data in the county tables were summarized to obtain totals for States; other State totals were run on the computer, and State cross-tabulations were run for farms having gross sales of $2,500 or more. Data for divisions, regions, and the United States as a whole were obtained almost entirely by summing the data from the State tabulations (more than 90 percent of the tables in volume II were posted from volume I tabulations plus data from publications of previous censuses) but some special computer runs were necessary for special frequency classifications of crops and livestock.

1964 Historical Data

The historical data needed for the diary and the county tables were created in two ways. The 1964 data for farms with $2,500 or more in gross sales were tallied in a special computer run. The data were printed out and the late corrections in the 1964 data, which had not been carried to the 1964 file previously, were carried to the computer tape for the 1969 tables. The all-farm historical data were transcribed from the 1964 volume I tabulations.

Final Disclosure Analysis

Federal laws governing census reports prohibit the publication of statistical data that reveal information furnished by individual respondents. In keeping with the provisions of this law, a thorough review was made of all statistical tables prior to publication to locate and prevent any potential disclosure of confidential information. Some of this review, referred to as disclosure analysis, was accomplished by one of the computer tabulation programs. However, since the conflicting interests in publishing as much of the data as possible and yet preventing disclosure of information about any individual operation presented major problems in the computer programming, most of the disclosure analysis was done by statisticians following broad general guidelines regarding what might constitute a disclosure. Figures were suppressed not only if they would, by themselves, be direct disclosures, but also to prevent the derivation by adding or subtracting of a subtotal or total that would disclose information about an individual agricultural operation.

The limitations set up in the guidelines for determining disclosed information were stricter for county tables than for tables for larger parts of States, or entire States, divisions, regions, or the United States as a whole. Statistics for larger areas were not considered as likely to allow identification of data for any one farm as statistics for the smaller areas would.

The established guidelines usually applied to cases of only one or two farms reporting an item. If more than two farms reported an item, the item was not considered as a disclosure of information unless the information to be published would reveal, by comparison of different tables, that one or two farms had 95 percent or more of the total. Exceptions were generally the larger specialized operations, such as poultry, feedlots, greenhouses and nurseries, and the raising of selected crops in areas where they were rare, any of which might easily identify a specific farm.

The number of farms reporting an item was not considered a disclosure; only the related information about the item was suppressed for publication. County reports were not published for counties with less than 10 farms, because of the many possibilities of disclosure.

Comparative data from the 1964 Census of Agriculture were published with several of the 1969 tables. Because 1964 tables containing summaries of data for farms with sales of $2,500 or more had not been published previously at the county level, they were also reviewed for disclosures.

Several of the tables contained the same information arranged according to a different classification, so that when it was necessary to suppress a figure in one table, it might also be necessary to delete it in another table. In the same manner, if an item was deleted from one county table, it was necessary to delete the same item from one or more other county tables.