

# CONTENTS

## CHAPTER 8. Processing the Data

(Page numbers here omit the chapter prefix, 8 – , which appears as part of the number of individual pages. The prefix indicates the location of the material in the final consolidated edition of the **Procedural History**.)

	Page
<b>Introduction</b> . . . . .	1
Basic Approach . . . . .	1
Specifications for Processing . . . . .	1
Computer Programming . . . . .	1
<b>Receipt and Check-In</b> . . . . .	2
Congressional Residence Allocation . . . . .	2
Central Files . . . . .	2
Breaker Sheet Preparation . . . . .	4
<b>Microfilming the 1970 Census Questionnaires</b> . . . . .	4
Camera Units . . . . .	4
Staging for 100-Percent Microfilming . . . . .	5
100-Percent Microfilming . . . . .	5
Sample Microfilming . . . . .	6
<b>FOSDIC Processing</b> . . . . .	7
Equipment . . . . .	7
Preparation for Processing . . . . .	9
100-Percent Processing . . . . .	9
Sample Processing . . . . .	9
Problems . . . . .	9
<b>Diary Review</b> . . . . .	10
Introduction . . . . .	10
100-Percent Diary Review . . . . .	10
100-Percent Diary Closeout . . . . .	12
Field Reinterview . . . . .	13
Block Statistics Diary Review . . . . .	13
Sample Diary Review . . . . .	14
Sample Diary Closeout . . . . .	15
<b>Sample Coding</b> . . . . .	15
Introduction . . . . .	15
Coding Procedures . . . . .	16
Quality Control . . . . .	19
Organizing and Staffing the Coding Operation . . . . .	21
Training . . . . .	21
Organizing the Coding Branches . . . . .	22
Production . . . . .	23
Distribution . . . . .	24
Problem Referral . . . . .	24
Rectification . . . . .	24

---

## Processing the Data — Continued

---

<b>Supplemental Forms and Coverage Improvement Procedures</b> . . . . .	24
Introduction. . . . .	24
Military and Maritime Personnel. . . . .	25
Overseas Travelers . . . . .	25
Americans Living Abroad . . . . .	25
Processing of Overseas and Military/Maritime Reports . . . . .	25
Late Receipts of Individual Census Reports and “Were You Counted?” Forms. . . . .	26
Post-Enumeration Post Office Check . . . . .	26
Introduction. . . . .	26
Jeffersonville Procedures. . . . .	27
National Vacancy Check. . . . .	30
Count Adjustment. . . . .	30
<b>Computer Processing</b> . . . . .	31
Equipment. . . . .	31
Staffing. . . . .	32
Computer Processing of 100-Percent Data . . . . .	32
Computer Processing of Sample Data . . . . .	35
Development of the Base File . . . . .	37
Fourth Count Processing. . . . .	37
Fifth Count Processing . . . . .	38
Sixth Count Processing . . . . .	38
Subfile Creation—Phase 123 . . . . .	38
Tabulations for Puerto Rico, Outlying Areas, and Americans Abroad . . . . .	38
Subject Reports . . . . .	38
Publication Processing . . . . .	39
Computer Time. . . . .	39
<b>Costs</b> . . . . .	40
<b>Bibliography</b> . . . . .	42

---

## Chapter 8. PROCESSING THE DATA

### INTRODUCTION

#### Basic Approach

Since the major purposes of collecting and processing the census returns are to determine the population counts needed for congressional apportionment and to produce tabulations of data required to carry out the Bureau's publication programs, the processing steps were geared to accomplish these purposes.

The major steps in processing the data from the 1970 census are outlined below. Steps 1-4 and step 7 were performed at the Bureau's Jeffersonville, Ind., facility; steps 5, 6, and 8 were accomplished on the electronic equipment at Bureau headquarters in Suitland, Md., and at other installations under census confidentiality restrictions, with referral to Jeffersonville for certain related clerical operations.

1. Receiving shipments from the census field offices containing the address registers for about 250,000 enumeration districts (ED's) and approximately 70 million household questionnaires, as well as other forms from which data would be extracted and incorporated in the census tabulations.

2. Checking in the questionnaires against control records.

3. Verifying the identification of each ED, so that the data would be allocated to the correct geographic area.

4. Microfilming the 100-percent data and exercising quality control of this operation.

5. Scanning the microfilm of the 100-percent data by FOSDIC 70 (Film Optical Sensing Device for Input to Computers) which "read" the marked circles, performed certain minor edits, and produced a detail tape of the data plus a printed-out diary for clerks in Jeffersonville to use in reviewing the counts. If the microfilm was unreadable, if the population or housing counts for an ED differed from the control counts, or if one or more questions were unanswered beyond acceptable tolerance limits, corrective measures were taken and a recycling through steps 4 and 5 was required.

6. Sorting, editing, and unduplicating the recycled ED's; tabulating the 100-percent data; resolving geographic problems; arranging the data in table format for output on the high-speed printer or for input to the Linotron photoelectric typesetting process; and pre-

paring the data for release on microfilm or in the form of summary tapes for census users.

7. Coding in FOSDIC-readable form the items on the sample questionnaires which were not coded by the respondent or the enumerator.

8. Following procedures similar to steps 4 through 6 for the sample questionnaires. In addition to the computer operations in step 6, the sample processing included weighting the 5-percent, 15-percent, and 20-percent sample data to 100-percent levels.

#### Specifications for Processing

Specifications for the processing originated in the Population and Housing Divisions, in the Systems Division for the operational aspects of the processing, and in the Data User Services Office and the Geography, Statistical Methods, and Publications Services Divisions for their areas of expertise.

The subject-matter divisions provided table outlines and detailed editing and coding specifications for both computer and clerical processing of the data. Tolerances for accepting data at the ED level were established jointly by the divisions having subject-matter or processing responsibility.

Most processing specifications were developed for the 1968 dress rehearsals of the 1970 census, although subsequent modifications were made in 1969 and 1970. Procedures for sample coding were refined in the winter and spring of 1969-70.

#### Computer Programming

Programming for the 1970 census differed from that in previous censuses in a number of ways, being affected particularly by the following developments during the 1960's: (1) The decision to use the mail-out/mail-back technique for a major part of the enumeration; (2) the expansion of geographic detail; (3) the expansion of demographic detail; (4) the introduction of the Linotron technique for photocomposition; and (5) the growth of the Bureau's role as a service agency.

As part of the mail census which covered more than 60 percent of the U.S. population, over 30 million residential addresses on magnetic tape were organized by computer into groupings relatable to census blocks, block groups, ED's, tracts, and larger areas. The tapes were used to generate address labels for questionnaire mailing pieces and to prepare over 100,000 address registers and directories for enumeration control. (For details, see chapter 3.) The population and housing

information collected in the census (e.g., identification of persons of Spanish heritage and of Indians by tribe, and assignment of place of work to the tract and block level of detail) contained data for determining the nature and extent of social trends and problems for specified geographic areas. These data had to be accessible to users on tape and capable of analysis by their computer programs, as well as in the traditional published form. (See chapter 13 for details.)

## RECEIPT AND CHECK-IN

Incoming shipments of completed questionnaires and other materials from the 399 district offices in the United States and Puerto Rico arrived at the Bureau's Jeffersonville facility between May 11 and September 22, 1970. In all, approximately 425 truckloads, with a gross weight of 6.1 million pounds, were received.

After shipments were unloaded and accounted for, the cartons were sent to the appropriate units for disposition. (See fig. A.) Address registers were delivered to central files for reference purposes, and supplemental forms (overseas census reports, crews-of-vessels questionnaires, etc.) were organized for processing.

The check-in operation, devoted to the household questionnaires and their control forms, was carried out in the following manner: Each district office's questionnaires were brought from the unloading area or from temporary storage to bays designated for check-in. Four groups of 15 to 20 clerks sorted the cartons of questionnaires by county and checked them against control lists. These were lists of all ED's in State, county, and ED order that existed prior to enumeration. To update these control lists, all population and housing count worksheets (form D-134) were reviewed for handwritten ED numbers after the ED cartons of completed questionnaires had been checked in. If a handwritten number on the form D-134 was not printed on a control list, the number was added. The cartons were placed in ED order on rolling bins or pallets and moved to another bay for completion of breaker sheets. (See below.) Check-in also involved the location of missing ED cartons and missing control lists.

The district offices routinely packed the questionnaires for a small ED in a plastic envelope and placed several of these envelopes into one ED shipping carton. As these questionnaires for small ED's occasionally were misplaced or combined with others during the next stages of processing, it was decided in August 1970, after processing was well under way, to have each ED placed in its own carton regardless of the number of questionnaires involved. After that date incoming ED's were separated and new cartons were assembled and labeled as necessary during check-in.

Because 1970-model bins were not available in the numbers needed for circulation during May and June 1970, and since there was a shortage of 1950- and 1960-model bins, pallets and bins without dollies were used to hold low-priority ED's until 1970-model bins were released at a later stage in processing. (The use of the 1960 bins, which were designed to hold materials of dimensions radically different from those used in 1970,

was particularly unsatisfactory; one-third of each 1970 ED carton protruded over the edge of the shelf, and the overlapping portion of the carton tended to collapse under its own weight.)

The lack of warehouse space for pallet storage and the fluctuating volume of material arriving forced relocation of some of the unloading operations and subsequent storage. The shortage of rolling bins added to the time and expense of check-in and of all of the subsequent steps in processing, inasmuch as the ED cartons frequently had to be transferred to and from the available bins, and the bins had to be lifted on and off the available dollies to keep materials moving in priority order.

## Congressional Residence Allocation

Each member of Congress had the option of being enumerated at his Washington-area address or in his home State. For Congressmen and Senators who made the latter choice, the appropriate population and housing data were tabulated for his home address. In this event housing information only was obtained for his Washington-area address and that housing unit was classified as "vacant--usual residence elsewhere."

## Central Files

In May 1970, rows of 9-foot-high shelves were set up in three bays to accommodate the incoming address registers, reference materials, and sample-coding supplies, and arrangements were made to provide controls over their use.

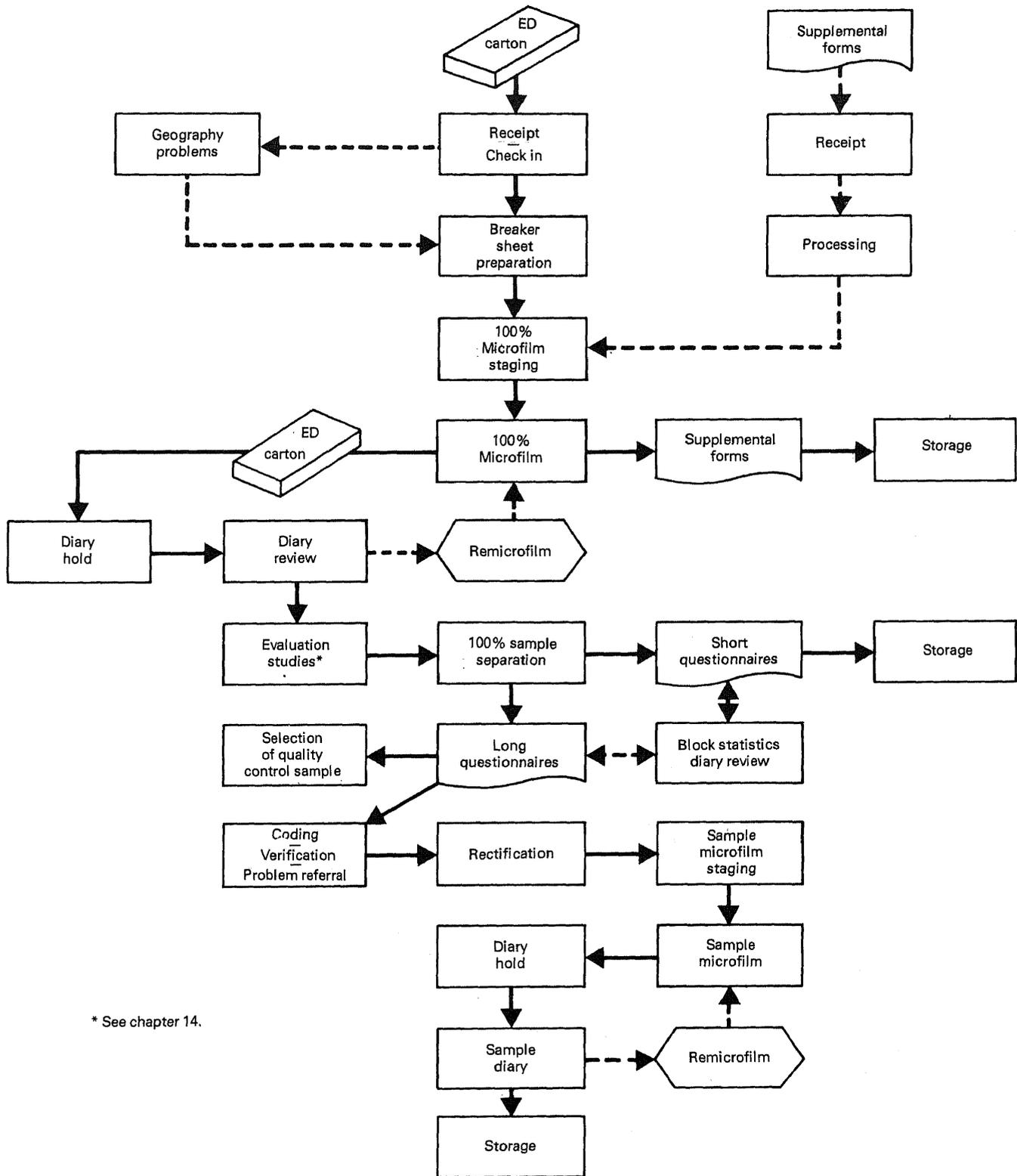
Shelf space was assigned for the address (ED) registers and directories for each State and county before their arrival. When actual shelving of the materials took place, it was necessary to set up "continuation sections" for many counties and to house overflow on a covered causeway between two buildings until shelf space could be reallocated. In January 1971, the central files were moved to another building and again shelved in complete State and county order.

Monthly requests for ED registers ranged from 21,000 to 70,000. Inasmuch as there frequently were as many as 11 different operations requiring these registers, a system of priorities was used in checking them out. These priorities varied from week to week, depending on the date when each operation had to be completed or the emphasis to be placed on one operation as against another. The recording and shelving system was changed several times to expedite the movement of registers to and from the people making the requests, and, at the same time, to account for each register.

An inventory was taken in March and April 1971 to reconcile the actual holdings of address registers in the central files and in each of the other operations where ED's were being processed with the ED's reported in the 1970 sample control listing.

A variety of reference materials also was stored and charged out of these central files. Among the holdings were 435 volumes of city street and business directories, 350 different ED directories (for identifying the ED's in

Figure A. Abbreviated Flow Chart of 1970 Census Processing in Jeffersonville, Ind.



\* See chapter 14.

Post Office city-delivery areas), 450 different tract and block directories (for assigning tract and block numbers to given addresses), three copies each of 460 block header record books (block header records were alphabetical and numerical listings of street names and house number ranges for which ED, tract, and census block numbers were provided), 750 national ZIP code directories, and numerous atlases and specialized directories.

An average of 27 persons, including one supervisor and one assistant supervisor, maintained the central file system.

### Breaker Sheet Preparation

To identify each ED in the computer, and to compare the population and housing counts reported for the ED by the district office with the totals obtained in processing the individual questionnaires, a FOSDIC-readable document, called a "breaker sheet," was completed for each ED. The breaker sheet preceded the questionnaires for each ED through microfilming.

The breaker sheet was prepared in three operations. Using computer tape prepared from the Bureau's master reference file, the high-speed printer overprinted approximately 250,000 breaker sheet forms with the State name and code, the county name and code, and the ED number. In addition, each ED was identified as "blocked" or "unblocked" (indicating that the questionnaires were or were not to be coded to census blocks) and as to whether the ED was one covered by a mail address coding guide. All codes were printed in a FOSDIC-readable format. The breaker sheets were printed, checked, and shipped to Jeffersonville on a flow basis between April 25 and July 16, 1970.

In Jeffersonville the breaker sheets, which arrived as continuous forms, were separated, packaged in plastic envelopes by county, and sent to the breaker-sheet preparation unit. Using the population and housing unit counts found on the control forms prepared in the district offices, a staff averaging 14 clerks entered these counts on the breaker sheets so that they could be "read" by the FOSDIC equipment. If a preprinted breaker sheet was found to be missing, dirty, or wrinkled, a new one was prepared manually by entering both the counts and the identification information required for FOSDIC reading.

The master reference file used to prepare the breaker sheets contained identifications for the ED's as they had been delineated prior to the enumeration. During the enumeration, and later in Jeffersonville, many ED's were split (divided into two or more parts) because of geographic problems or because the ED contained too many people or housing units to process as one work unit. No ED could be processed if it contained more than 9,999 persons or more than 999 housing units. To split an ED, the questionnaires were separated and the counts determined for the newly created ED's. For each of these ED's the original breaker sheet was used for the original ED and a new breaker sheet was prepared by hand for each split, with an alphabetic suffix "A" added to the original ED number, and "B," "C," etc., added to the ED number for each split-off portion.

Geographic errors found during the enumeration required correction of geographic codes and/or movement

of questionnaires from one ED carton to another. Many of these corrections were made during the clerical preparation of the breaker sheets. Cases which could not be resolved by the clerks were referred for determination to a group of geographic specialists located in the preparation area.

To ensure an acceptable level of quality in the preparation of breaker sheets, the identification items on all hand-prepared breaker sheets were verified 100-percent, and the population and housing unit count entries were verified on a 10-percent sample of all the sheets. After the clerks gained experience, all verification was confined to a small sample. Serious problems were caused when erroneous identifications were entered by hand. If a digit was not marked, the microfilm for the ED was rejected in the FOSDIC process without transferring any of the data to magnetic tape. The breaker sheet for the rejected ED then had to be corrected and the entire ED remicrofilmed.

## MICROFILMING THE 1970 CENSUS QUESTIONNAIRES

### Camera Units

Between April 1969 and May 1970, 40 automatic microfilming camera units were installed at the Bureau's facility in Jeffersonville. Each camera unit consisted of a 16mm camera, a document-transport mechanism, a page-turning arm, two joggers, and a light meter. The camera, document-transport mechanism, and page-turning arm were connected electronically so that the operation of these three pieces of equipment could be synchronized.

The camera was mounted on an "elevator" shaft which permitted vertical movement of the camera. Controls were built into the equipment which automatically stopped the camera at a predetermined point at both the lower and upper end of the elevator shaft. The camera used 200-foot reels of 16mm microfilm, allowing exposure of a maximum of 1,870 frames per reel at the rate of 60 to 100 frames per minute.

The document-transport mechanism consisted of a feeder which took the document (the questionnaire) from a hopper to a vacuum belt. The belt transported the document into position for filming, held it there, and then moved the document to a stacker when the filming was completed. The stacker took the document from the belt and stacked it with others for return to its original carton.

The page-turning arm consisted of a revolving arm which was engaged to pass over the multipage sample questionnaires, picking up and turning pages as it did so.

Two joggers, which were separate from the camera and the document-transport mechanism, were used to jog the questionnaires until their edges were aligned.

Each camera unit was in an individual cubicle, enclosed on three sides with black drapes, in the air-conditioned camera bay.

Each camera unit was staffed by a camera operator and one or two helpers. In "100-percent" microfilming (see below), one helper assisted in preparing the questionnaires to be filmed and another in stacking the filmed questionnaires. In sample microfilming one helper performed both of these functions.

**Testing.**--Each time a camera was to be used, it was first subjected to a series of mechanical tests, usually conducted daily. The results of these tests were analyzed to determine if any modifications or adjustments were required on the machines and to set the operating conditions for obtaining optimum production from each camera. The tests, in sequence, were as follows:

1. An illumination test, in which light-meter readings were taken from three different positions on the questionnaire under the camera. A maximum deviation among these readings of 20 percent was allowed.

2. A step test, in which camera exposures were made under gradually increasing intensities of light, and the developed film compared with GSA (General Services Administration) Federal Standard gray scale chips to determine density and reflectance. A density value of 1.05 was the accepted standard for light-meter readings between 18 and 48 lumens; the acceptable density range, once the 1.05 value had been obtained, was 0.85 to 1.25.

3. A dip test, in which the camera functions were measured in terms of film feed, shutter action, and focusing. The standards required the following results:

- a. Tilt in the questionnaire image on the film resulting in no more than 1/8-inch difference between two horizontal corners, as measured from the edge of the film at a reduction of 1:2; and

- b. Maximum vertical shift of the questionnaire to the edge of the film of 1/16 inch in either direction as measured at the 1:2 reduction ratio. The absolute minimum space allowed was 1/4 inch between frames at this ratio. Less than this amount of space would have resulted in the loss of exposures, and excessive space would have caused waste of film.

4. A production simulation test, in which a camera rate of at least 50 exposures per minute was achieved, with a minimum amount of mutilation and exposure failures.

5. A sustaining test, in which the same materials were microfilmed several times on each of two cameras and the results compared to check any variations among camera runs and between different cameras.

6. A latent-image test, in which 28 film strips, all exposed in one camera run, were developed at varying intervals (e.g., the first strip 30 minutes after exposure, the second strip 1 hour after exposure, etc., up to 50 hours for the 28th strip) to check the quality of the resultant product.

Approximately 40,000 blank household questionnaires, otherwise acceptable for census processing, were coded in a prescribed manner by Bureau clerks so that documents of a known, controllable quality could be used in testing the cameras.

**Camera maintenance.**--A team of 10 technicians was employed to maintain the 40 automatic microfilm cameras. These technicians, who familiarized themselves with the cameras during the microfilming operations, were directly responsible for camera modifications that improved overall performance and production capabilities. Equipment "down time" ran higher than anticipated, due mostly to broken document-transport belts, but this situation was corrected by the installation of heavier belts.

### Staging for 100-Percent Microfilming

This operation received questionnaires each day from the check-in unit on a flow basis. Clerks opened each ED carton and prepared the questionnaires for microfilming of the 100-percent data by arranging them face up, each open to pages 2 and 3, with the sample questionnaires first. If the questionnaires in the carton were for more than one ED, separate ED cartons were prepared as necessary. The appropriate breaker sheet then was placed on top of the questionnaires in each carton. (All of the breaker sheets in the same county for ED's with zero population and housing units were placed in one carton.) Carton identification was checked and corrected as required, and the cartons were placed in rolling bins by county and in ED sequence within county. A set of bin tags was fastened to the bin with the "Ready for Camera" tag visible.

Bins of questionnaires ready for microfilming were arranged in rows adjacent to the camera room according to State priority. Requirements for camera operations were approximately 2,000 ED's per 40-camera shift, or 50 bins (with an average of 40 ED's to a 1970-model bin).

Staging required a staff which averaged 65 clerks in any 1 week; however, a large amount of overtime was necessary to meet the target dates and maintain the appropriate backlog of work to be microfilmed. Full-time staging operations for 100-percent microfilming drew to a close during December 1970.

The following sections describe the microfilming operations, including the staging and filming of the sample questionnaires which was a similar but later step in processing. (See flow chart, fig. A.) Microfilming of supplemental forms (individual census reports, overseas census reports, etc.) is not described here; these were microfilmed routinely as a matter of record after they had been processed and before they were stored.

### 100-Percent Microfilming

This operation was so called because it entailed microfilming of that population and housing data on the household questionnaires which had been collected on a 100-percent basis. These data appeared on pages 2 and 3 on both the short and long (sample) questionnaires.

The number of household questionnaires per ED ranged from zero to 800 or 900, with 20 percent being sample questionnaires. There were approximately 250,000 ED's and 70 million questionnaires in all, representing the enumeration of 3,141 counties or county equivalents in the United States.

In the 100-percent microfilming operation the "breaker sheets" and the questionnaires for that ED, open to pages 2 and 3, were fed through the camera unit. The two pages were photographed simultaneously on one microfilm frame; the questionnaire then was moved on and another questionnaire was moved into place under the camera lens. The frames for each roll of film were preceded and followed by frames containing a photograph of the identification board. This board displayed the camera work-unit number (a serial number consisting of the camera identification number and the microfilm roll number) and the date of microfilming. Several ED's usually were photographed on one roll of film; the ED's sometimes were selected out of ED sequence within a county in order to fill a roll of film, but in general no ED was split between two rolls.

After microfilming each exposed roll was developed by a private contractor in the film processing room which adjoined the camera area. If the roll passed development and density checks, it then was shipped to Suitland for processing on FOSDIC. (See pp. 7-10.) Any roll which failed either of these checks was re-microfilmed in its entirety.

One shift supervisor was assigned full-time to the camera room in December 1969, and a film evaluator was assigned in March 1970 along with the first two camera unit supervisors. Six additional unit supervisors were added on April 5, and the first 19 helpers were brought in 2 weeks later. The supervisors trained these helpers to be camera operators over the next 2-week period. Meanwhile, additional helpers were hired; these, in turn, were assigned to the camera operators who trained the new helpers to be operators as well. This sequence continued until an adequate number of operators was qualified. The original supervisory staff received 8 hours of administrative training; supervisory replacement was accomplished by promoting qualified camera operators.

Ultimately, a total of 419 employees were hired for the microfilming operation. There was a high attrition rate among the camera crews, due in part to the departure of students who were employed only during their summer vacation and in part to the long workdays necessary to complete the 100-percent microfilming operation in time to meet deadlines. Actual microfilming began in late April 1970, with an 8-hour day shift. An 8-hour night shift was added on June 1. On July 11 a standard 6-day week was established by making Saturday a mandatory overtime day. On August 17, 1970, the shifts were changed to 10 hours a day each, Monday through Friday, plus two 8-hour shifts on Saturdays. Attempts were made to staff half of the cameras on Sundays with volunteers, but this practice soon was discontinued because there were insufficient volunteers. When the work week consisted of two 58-hour shifts daily production ranged from 1-1/4 to 1-1/2 million exposures per day, but high absenteeism during this period (August and September 1970) required the recruiting and training of Bureau clerks as reserve camera crew members.

Work on 100-percent microfilming was substantially completed by early November, when approximately 65,000 rolls of microfilm and a total of 102 million exposures had been made (including remicrofilming). Night work and overtime then were suspended until needed again for sample microfilming (see below), but 100-percent microfilming continued as needed for Puerto Rico questionnaires, for remicrofilming, etc., until the end of June 1971. Approximately 68,000 rolls (105 million exposures) were involved in this operation.

After microfilming, the cartons of questionnaires were sent to the "diary hold" area, where they were kept until a review of the FOSDIC diaries (see p. 10) indicated that they either could be released for sample coding or required remicrofilming. Beginning in January 1972, accepted microfilm of 100-percent data was spliced into 3,614 master reels averaging 2,200 feet per reel, in county and ED order, and stored. Splicing was completed in December 1972, and required 1,450 man-days. Approximately 2 million feet of microfilm, containing superseded data for about 60,000 ED's, were eliminated. (Beginning in July 1972, the 100-percent microfilm was duplicated for the Bureau files. The project was completed on March 30, 1973.)

After release, the 100-percent questionnaires were separated from the sample questionnaires and placed in special storage cartons by county. Blocked and non-blocked ED's were so marked in case the blocked ED's would be needed again during the block statistics diary review. (See pp. 13-14.) (Blocked ED's were those in which data were tabulated by census block; in non-blocked ED's the ED was the smallest tabulation area.) The cartons then were placed on pallets for transfer to storage areas.

The sample questionnaires and breaker sheets were placed in freshly labelled cartons, with no more than one ED each, and were sent to a "diary hold" area until needed for coding. (Questionnaires were not microfilmed for sample data until they had been individually coded by clerks. The coding operation is described on pp. 15-23.)

### Sample Microfilming

This operation entailed microfilming the completed pages of the 15-percent and 5-percent sample questionnaires. Although pages 2 and 3 of each sample questionnaire contained 100-percent data that already had been microfilmed and tabulated, it now was necessary to microfilm these pages again along with the coded sample pages so that all the data for the sample housing unit or sample group-quarters persons would appear on the same record.

Sample microfilming encompassed the same ED's as were processed through 100-percent operations; there were about 13.7 million sample questionnaires. These were staged in a manner similar to that for 100-percent microfilming. (See above.) The breaker sheet and the filled-out pages of the sample questionnaires in an ED were filmed, two pages at a time on one frame. The operation differed from 100-percent microfilming in the following ways:

## FOSDIC PROCESSING

1. The page-turning device was used. The page-turner, synchronized with the camera shutter, continued to turn the pages of a questionnaire until a photoelectric cell sensed a mark on the questionnaire which indicated that there were no further data on the remaining pages. At this point the document-transport mechanism was activated and a fresh questionnaire was brought into position under the camera lens while the document just completed was carried off to the stacker. The page-turner then repeated its action. Each camera operator was given 3 to 5 days of training in the operation of the page-turning arms.

2. One helper was sufficient to assist the camera operator. The helper removed the ED cartons from the bins and jogged and loaded the questionnaires into the hopper for microfilming. He removed the microfilmed questionnaires from the stacking unit, jogged them, placed the questionnaires back into their original boxes, and reloaded the boxes onto a bin. He also repaired mutilated questionnaires as necessary during the camera operation and positioned them for microfilming. Where feasible he assisted the operator in recordkeeping and in actual operation of the camera.

The timing and workload for sample microfilming depended on the speed with which the sample questionnaires were hand-coded by clerks in the coding units. The first sample questionnaires were microfilmed on October 22, 1970, but, due to delays in releasing ED's for coding, it was necessary to curtail microfilming for several months. This necessitated retraining some of the operators and helpers when quantity production resumed in March 1971 on a two-shift basis. Sample microfilming was a slower operation than 100-percent microfilming, because each questionnaire had to have a number of pages turned and thus remained under the camera for a longer time. However, double-shift output soon reached 650,000 exposures per day, and in April a steady production rate of 1 million exposures a day was attained with about 180 employees on two 8-hour shifts.

The following factors contributed to this high rate of production: There was a sharp decrease in "down" time on the cameras, as most of the mechanical problems encountered in 100-percent microfilming had been remedied. The questionnaires were received in better condition than anticipated, so that little time needed to be spent in repair work during microfilming. Early in the operation paper clips and other foreign objects had caused problems, but these were eliminated by the establishment of a premicrofilm carton inspection. New work standards and an incentive pay plan were put into effect during the first week of April. This last factor tended to reduce personnel turnover but placed emphasis on quantity rather than quality of production, and this played a part in the increased amount of remicrofilming.

Sample microfilming (including remicrofilming, processing for Puerto Rico, etc.) was completed in early October 1971. A total of 85,659,515 exposures were made on 54,612 rolls of film. These totals included approximately 12.5 percent due to remicrofilming work.

### Equipment

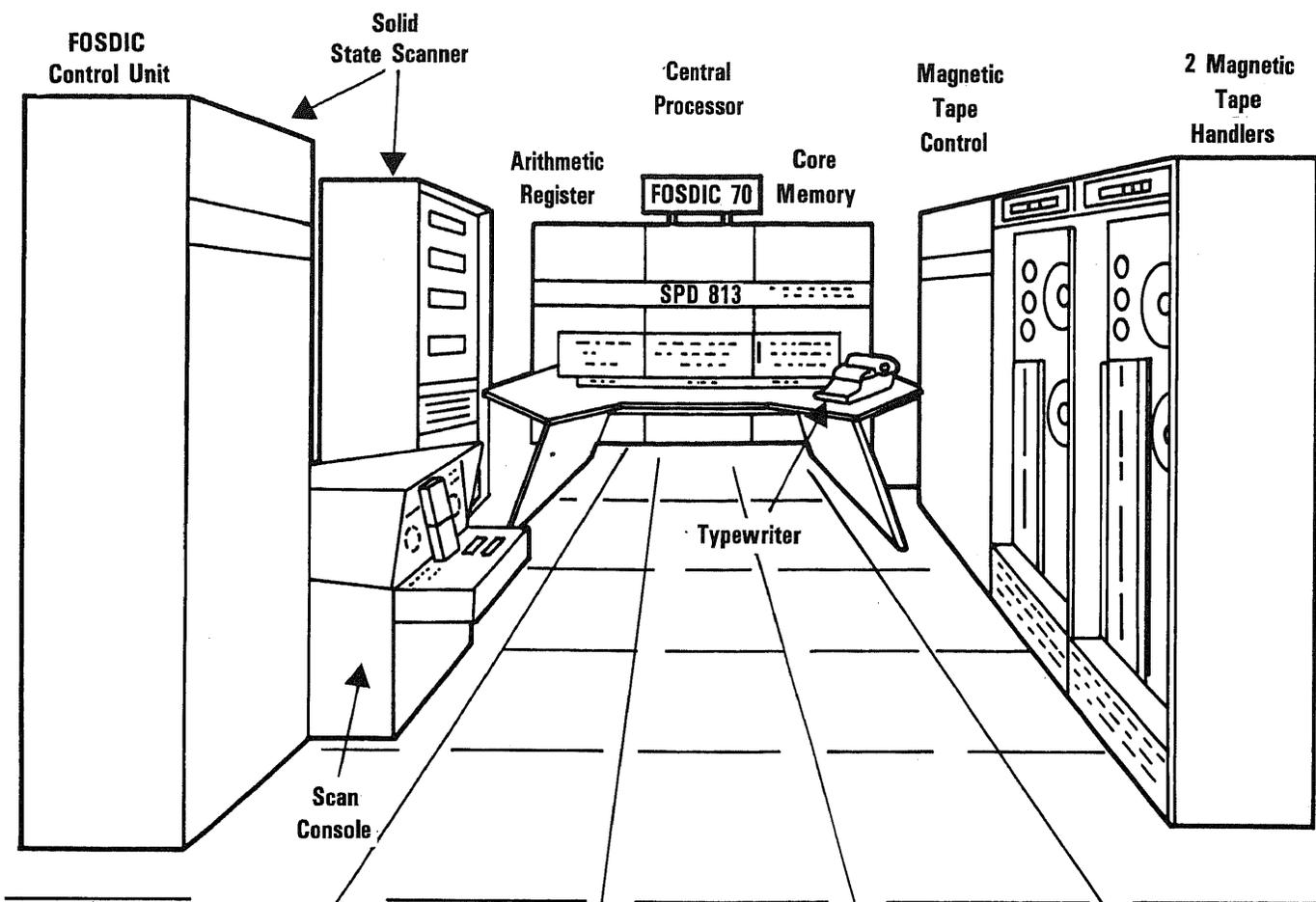
Part of the 1970 census processing involved the use of a high-speed electronic device called FOSDIC (Film Optical Sensing Device for Input to Computers). This device reads information from the microfilm negative of the specially printed and marked census questionnaires and transfers the data to magnetic tape which can then be processed by computer. FOSDIC, using a moving beam of light, first scans the microfilm for certain index marks on each questionnaire image and checks for image alignment and proper photographic reproduction. It then scans a series of blank circles printed on the questionnaire margin. This step, called "setting the threshold," tells the machine what a blank circle looks like. The beam finally scans the questionnaire image, picks up as answers any circles a certain amount lighter on the negative film than the blank circles, ignores any other marks, and translates these answers into Binary Coded Decimal Excess-3 computer language on magnetic tape for input to the Bureau's computers.

FOSDIC was developed jointly by the National Bureau of Standards and the Bureau of the Census during the 1950's for processing the 1960 decennial census data. The original FOSDIC equipment held a single 100-foot roll of 16 mm. negative microfilm which was scanned by a cathode-ray tube beam. The device used vacuum tubes and an external plugboard for programming. It had one tape unit, a one-digit memory, and processed about 100 microfilm frames a minute. In the 1960 census approximately 50,000 100-foot rolls of microfilm were processed on four machines.

It was known that the 1970 census would involve processing about five times this amount of film (because of the adoption of individual household questionnaires for 1970 instead of schedules containing data for four households on one sheet, as had been used for much of the 1960 enumeration.) If the time needed for processing remained the same, at least 20 of the 1960-type scanners would be needed to complete the work within the required time period. During the decade between 1960 and 1970, however, a number of refinements were made in the FOSDIC system. The equipment used for a 1970 census test in 1967 was a modified 1960 tube-type machine designated as SPD 413 (Systems Peripheral Device); the designation derived from the fact that it could hold 4,000 four-digit words in its core memory, it had one arithmetic register, and three input-output channels. It had one-half the memory and about one-half the scan speed of the 1970 system.

The version of FOSDIC ultimately developed for the 1970 census (see fig. B for illustration), called FOSDIC 70 (SPD 813), had a core memory with a capacity of 8,000 four-digit words, one arithmetic register, and three input-output channels. FOSDIC 70 was solid-state (rather than utilizing vacuum tubes), had two tape units, and held a 220-foot roll of microfilm, thus reducing the ratio of preparatory time to operating time. It scanned the 1970 questionnaires at an absolute pass-through rate of 750 frames per minute and an effective rate of approximately 200 frames per minute (the average production rate for an 8-hour shift). Six of these machines were built for

Figure B. The FOSDIC 70 System



1970 and were ready to begin the FOSDIC processing operation just as the first roll of film arrived. Before a machine was considered ready for production an engineer's test was conducted on it to ensure that the machine's adjustments were correct. A roll of microfilm containing exposures of questionnaires with known markings was processed through each machine. If the number of pickups (blanks read as answers), drops (answers not read), or answer differences (answers recorded wrong) was greater than approximately 0.005 percent, the machine was taken out of service until the required adjustments were made. This test was repeated on each machine prior to the start of every shift during the entire production operation.

There were no operational tests of the FOSDIC 70 system before the census processing began; that is, 1970 census programs, such as used in the dress rehearsals, had never been used to process "live" data on FOSDIC 70.

During production an automatic typewriter provided the FOSDIC operator with information on record counts and processing problems, such as the rejection of ED's with incomplete breaker sheets or ED's processed with impossible State codes. The typewriter printed out the State, county, and ED numbers, the total number of output records, and a list of the acceptable ED's in the work unit. (A FOSDIC work unit consisted of the microfilmed

questionnaires for 30 to 40 ED's.) As the microfilm for an ED was processed through the FOSDIC operation, the number of blanks and possible answers was tallied for each item. Likewise, each person and each housing unit was tallied.

During 100-percent processing (see below) FOSDIC produced two magnetic tapes: (1) the detail tape, a record of everything "read" by FOSDIC, used in computer processing, and (2) the diary tape, which contained population and housing counts for each ED in the work unit and other information to be used in clerical review.

During sample processing only the detail tape was produced by FOSDIC. This tape was similar to the 100-percent detail tape. A diary was produced in subsequent computer processing and editing; each diary covered 14 sample FOSDIC work units.<sup>1</sup>

The detail tapes produced in the FOSDIC operation were consolidated by computer into fewer units, known as "sort" tapes. This sorting process was monitored by the production of a message giving the number of ED's read (and isolating those not read), so that ED's or entire FOSDIC tapes that might have been misprocessed could be detected in the diary review. (See pp. 10-12.)

<sup>1</sup>A sample work unit contained 13,000 to 14,000 microfilm exposures (but no more than 20 rolls of film), or the sample questionnaires for about 35 ED's.

## Preparation for Processing

After the microfilm was developed and had passed inspection in the Bureau's Jeffersonville facility, it was shipped by air to Bureau headquarters in Suitland. Here the microfilm was to be stored until sufficient ED's for one State had been received. When rolls were assembled for FOSDIC processing, the roll numbers of the microfilm were listed in ascending roll-number order on a FIT (film issue transmittal) which had been preprinted with the State name, State code, and FOSDIC work unit number within each State. The microfilm and the FIT then were sent for processing through the FOSDIC system.

## 100-Percent Processing

FOSDIC 70 was programmed to perform these basic tasks:

1. Scan each breaker sheet and questionnaire page for clerical or respondent entries. Unless the breaker sheet contained certain identification entries, the entire ED was rejected at this point. If the breaker sheet was accepted, the FOSDIC scanned the individual questionnaire pages separately for responses.

For page 2, the FOSDIC scanned the last "person" line and each higher line until it found an entry in either sex (item 3) or race (item 4). Once it found an entry in either item for a given line, it scanned all entries for this line and all lines above it. For a given line to qualify as representing a person, it needed two or more entries from the total possible entries. Decade and year of birth (item 7) were treated as one entry. Month (item 6) was not counted as a qualifying entry, although it was read.

Page 3 was scanned for all possible entries; individually scanned pages were then put together to form complete questionnaires.

2. Impute data for missing persons, determine occupancy status of housing units, and perform other minor edits for consistency in order to prepare the data for the next phase of processing.

3. Print out a diary (a listing for clerical review) for each ED showing (a) the final counts of population and vacant and occupied housing units, together with the percent of difference between these counts and the counts on the breaker sheet, and (b) for each item the percent of entries that should have been completed but were left blank, could not be read, or were edited in the FOSDIC operation.

4. Prepare a multiframe tape (each ED being a file) to facilitate identification of the file for a particular ED on the computer tape should reading problems develop during subsequent computer processing.

5. Provide printed-out control data about the microfilm, including the number of ED's dropped (i.e., rejected because of unacceptable breaker sheets or not read because of defective filming) and a list of the ED's retained.

FOSDIC processing of the 100-percent data on all of the household questionnaires began in May 1970. State counts for apportionment were completed by November 1970, and detail counts were completed in December with the six machines being operated most of the time in three 8-hour shifts a day, 6 days a week. Production per machine averaged over 275,000 frames a day. In all, 77,319 rolls of microfilm were scanned (some had to be scanned more than once) during 100-percent processing, with a weekly rate peaking at 6,476 rolls. With a base of 250,000 ED's, 253,509 ED's were received from microfilming (some having been reprocessed and recounted through that operation); 64,776 were returned for remicrofilming, so that a total of 318,285 ED's were put through the FOSDIC operation. Of these, 71,354 ED's, or 28.6 percent of the base, had to be reworked.

## Sample Processing

Except for counting blank and unreadable pages, none of the determination actions used for the 100-percent data were taken when the microfilm of the sample data was processed through FOSDIC 70. The sample data were simply transcribed to magnetic tape as found, and review and clerical repair took place after computer editing of the tape. As in the 100-percent operation, this process consolidated a number of rolls of microfilm into a single output tape and prepared the output in a multiframe format (each ED a separate file) to assist in identification of ED's if tape reading problems occurred.

FOSDIC sample processing began in late January 1971, with four machines (two machines had been removed for shipment to Canada for use in that nation's census) operating two shifts a day, 5 days a week. Between this time and completion in early October 1971, 72,855 rolls of microfilm were handled, including those for Puerto Rico. Although production reached 3,787 rolls in one week, maximum output was hindered by a sporadic flow of film from Jeffersonville (caused by backlogs in coding the sample questionnaires before they could be microfilmed). Of the 353,234 ED's processed (including those which had to be remicrofilmed), 106,303 ED's, or 30 percent, had to be reworked in the FOSDIC operation.

## Problems

As indicated, approximately 29 percent of the 100-percent microfilm rolls and 30 percent of the sample microfilm had to be reprocessed through the FOSDIC operation. This was necessary because of missing data, improper microfilm, FOSDIC malfunctions, lost or miscombined computer tapes, and the like. (The number of ED's requiring FOSDIC rework was not tallied by type of problem.) Post-census examination and laboratory analysis of a small sample of the work indicated that virtually all of the variation in registration appeared to be caused by mark reflectance, i.e., the blackness and coverage of the pencil marks made by respondents and Bureau coders. Coders' markings were significantly poorer than respondents', as evidenced by the need during sample processing to send some questionnaires to clerical coding units in the diary review operation (see below) to have coding circles remarked. Some FOSDIC reading problems also might have been attributable to questionnaires printed on paper that was deficient in opacity or

had too much surface brightness, or on which the index marks were too lightly printed.

## DIARY REVIEW

### Introduction

The purpose of the diary review operation was to find and correct suspected errors above certain tolerance levels in the 100-percent and sample data contained in the basic data tapes produced by the FOSDIC operation. Diary review was concerned with (1) systems errors, i.e., those errors which caused information to be lost or incorrectly recorded on the magnetic tapes, and (2) a minimal number of potential data errors, such as an excessive number of entries recorded as "other" for color or race (questionnaire item 4). When a tolerance level was exceeded, an alphabetic flag appeared on the diary printout. The printout was a summary of the data for each ED in a FOSDIC work unit. The 100-percent data for each ED had to be accepted before the ED was released for coding of the sample questionnaires. Every established ED, including those for crews of vessels, and every ED split for control purposes or for geographic reasons had to be accounted for to complete the tabulations for a State.

In the 100-percent phase of processing, pages 2 and 3 (containing the complete-count population and housing data) of every household questionnaire were microfilmed and processed by computer to produce the official counts of population and housing units in the United States as of April 1, 1970. The long questionnaires containing the sample data then were separated from the short questionnaires and passed through three specialized coding operations. (See pp. 16-18.) In these coding operations written responses to various questions were translated into marked circles which were "read" in the subsequent sample FOSDIC operation and converted to magnetic tape after the coded long questionnaires had been microfilmed. As part of the processing of the tape through the FOSDIC and subsequent computer operations, diaries--similar to those used in the 100-percent diary review--were produced for use in reviewing the sample data.

The diary review operation began with the 100-percent data in May 1970. This phase was completed in late October 1970; diaries for sample data were reviewed between February and October 1971.

In the 1960 census both 100-percent and sample diaries were produced during the initial computer processing of the basic record tapes. In 1967 the first 100-percent diary was produced directly from FOSDIC output, using the data gathered in a census pretest conducted in the New Haven, Conn., SMSA (standard metropolitan statistical area). This diary contained the population and housing unit counts, alphabetic reject codes, and other information concerning the 100-percent data items on a one-line printout for each ED. Processing of the New Haven data was similar to the 1970 census processing, with the basic data produced on one tape and the diary on another. The major difference was that a reject code on the New Haven diary caused erasure of data from the data tape. If, during the diary review, it was discovered that a reject flag had been caused by missing data and therefore could not be corrected, or by an error in control

counts, the ED had to be reprocessed through the FOSDIC operation using a bypass code to allow completion.

In August 1968 another modified version of the 1960 FOSDIC was used to produce 100-percent diaries for the dress rehearsals of the 1970 census conducted in that year. The dress rehearsal diaries contained six lines of information for each ED, with 73 100-percent data entries possible, nine error flags, and five reject flags. The error flags, each of which covered one or more items, were used to indicate that a problem appeared to exist which should be investigated, but the flag did not prevent further processing of the data as it had in the New Haven pretest. Rejection flags were used primarily for problems that were caused by improper microfilming or incomplete or incorrect breaker sheets. For several of the error flags the serial numbers of the questionnaires responsible were also printed in the diary. Review of the dress rehearsal diaries and the corrections required led to the setting of tolerances for individual items rather than for groups of items. These tolerances by item appeared as parameters in the FOSDIC program for 1970, so that any one, or all, of the tolerances could be changed during the census processing as necessary.

Except for tolerances and some rearrangement of items, the 1970 census 100-percent diary was similar to that used in the dress rehearsals. Information concerning several items that was not needed for clerical review was blacked out on the original diary printout form but did appear on carbon copies of the printout used by subject-matter analysts.

### 100-Percent Diary Review

This diary operation began in the same clerical unit in Suitland which prepared the microfilm for processing on FOSDIC 70. After processing, the microfilm and its FIT were returned with a diary and a copy of the FOSDIC typewriter printout which listed by number each of the ED's processed in the FOSDIC work unit.

The count of the number of ED's on the FIT was compared with the same total on the FOSDIC typewriter printout. Following clerical review and adjustment of any differences, the FIT and one copy of the diary were placed in a plastic bag and shipped to Jeffersonville.

A diary shipment log was maintained in Suitland for each FOSDIC work unit. This log indicated when the FIT and typewriter printouts were received, when microfilm rolls were returned for reprocessing, when diaries were received, and when these diaries were sent to Jeffersonville.

The space allocated in Jeffersonville for the "diary hold" area was adequate for storage of about 400 bins of ED's--the production of 4 to 5 days of microfilming that would be awaiting diary review. The assumption that production work could be processed through diary review within 5 days of microfilming was unrealistic. The actual elapsed time between microfilming and receiving the FOSDIC-processed work-unit diaries in Jeffersonville was frequently 10 days or more. This delay, plus the time required to review the diaries, quickly caused the hold area to fill with bins of ED cartons, so that it was necessary to stack the bins until an estimated 1,200 bins--

stacked three high, holding questionnaires for about 60,000 ED's--were in "hold." More search personnel were necessary to locate and retrieve ED cartons. The height of the bins required ladders in the aisles and made it difficult to maneuver the forklifts necessary to hoist or lower bins from the upper levels. The arrangement of materials in the hold area on bins and shelving made it difficult to locate ED's needed in the diary review operation or to release those ED's that had been cleared. Much clerical time was consumed in waiting for the location of "lost" ED cartons.

Diary review was conducted by a staff of 120 to 220 persons. Their training consisted primarily of the reading of a procedures manual and demonstration of the various materials used; flow charts and graphs were used to explain the process. Supervisors and clerical personnel were expected to gain specialized knowledge through on-the-job experience.

Diaries were checked in by entering the FOSDIC work-unit number opposite the ED number on the control list completed at the time the questionnaires were received from the field. (See p. 2.) After all the ED's in a work unit were checked in, the diary for the unit was screened for reject flags. Questionnaires for ED's requiring review were ordered from the diary hold area by circling the ED numbers on the work unit's FIT; the matching address registers, if needed, were requested from the central files. (Originally, address registers were to be ordered for entire States, but the diary review work area was too small to hold the number needed and the registers also were in demand for other census operations. Priorities for the use of the address registers were established in mid-July 1970 and were revised frequently to coincide with processing deadlines.) When all of the required ED cartons and address registers for a given work unit were received, they were assigned to a review clerk.

All ED's not designated on the FIT for diary review were released for sample coding. In the beginning, the FIT accompanying each work unit had to be used in the diary hold area both to pull ED's for diary review and to release others for the sample coding operation. Later, FIT's were transmitted in duplicate so that one copy could be used exclusively for pulling ED's for review. Because of the time needed for search, receipts from storage continued to be sporadic, so that in order to meet diary review deadlines approximately 100 review clerks had to be added to the original staff of 120 and there was overtime work for about 6 weeks.

Review consisted of (1) identifying and examining the individual questionnaires, breaker sheets, and address registers that caused the appearance of flags on the diary; (2) editing, correcting, or remarking the original questionnaires and other materials where necessary; (3) referring unsolved cases to subject-matter or systems specialists; and (4) ordering remicrofilming and/or FOSDIC reprocessing where needed.

For 1970 there were nine different flags used to identify questionable 100-percent data in an ED. These flags were designated alphabetically as A, B, C, D, E, F, G, H, I, and K. (To avoid confusion with the letter J, J was not used.) The significance of each, together with the

statistical tolerances set, are shown below. In August 1970 certain tolerances were lowered to reduce the diary review workload in order to meet the deadline for apportionment counts and to speed the movement of ED's to sample coding; after evaluation of early tabulations it was believed this could be done without materially lessening the quality of the data.

- | <u>Flag</u>     | <u>Condition</u>  |                  |                       |                  |                   |   |                      |     |  |   |     |    |     |   |               |     |  |    |                 |     |  |    |               |    |  |   |                |     |  |
|-----------------|---|------------------|-----------------------|------------------|-------------------|---|----------------------|-----|--|---|-----|----|-----|---|---------------|-----|--|----|-----------------|-----|--|----|---------------|----|--|---|----------------|-----|--|
| A               | 1. The absolute difference between the population count from the questionnaires and the population count on the breaker sheet was more than 50, and the percent difference was more than 5,<br><u>or</u><br>2. The percent difference was more than 89,<br><u>or</u><br>3. The breaker sheet population count was missing.  |                  |                       |                  |                   |   |                      |     |  |   |     |    |     |   |               |     |  |    |                 |     |  |    |               |    |  |   |                |     |  |
| B               | 1. The absolute difference between the occupied and vacant housing unit counts from the questionnaires and the same counts on the breaker sheet was more than 20 and the percent difference was more than 5,<br><u>or</u><br>2. The percent difference was more than 89,<br><u>or</u><br>3. The breaker sheet housing unit count was missing.   |                  |                       |                  |                   |   |                      |     |  |   |     |    |     |   |               |     |  |    |                 |     |  |    |               |    |  |   |                |     |  |
| C               | The blank rate for the population items listed below exceeded the allowed tolerance:  |                  |                       |                  |                   |   |                      |     |  |   |     |    |     |   |               |     |  |    |                 |     |  |    |               |    |  |   |                |     |  |
|                 | <table border="0" style="width: 100%;"> <thead> <tr> <th style="text-align: left;"><u>Item No.</u></th> <th style="text-align: left;"><u>Characteristic</u></th> <th style="text-align: left;"><u>Beginning</u></th> <th style="text-align: left;"><u>Changed to</u></th> </tr> </thead> <tbody> <tr> <td>2</td> <td>Relationship to head</td> <td>10%</td> <td></td> </tr> <tr> <td>3</td> <td>Sex</td> <td>5%</td> <td>10%</td> </tr> <tr> <td>4</td> <td>Color or race</td> <td>15%</td> <td></td> </tr> <tr> <td>7D</td> <td>Decade of birth</td> <td>15%</td> <td></td> </tr> <tr> <td>7Y</td> <td>Year of birth</td> <td>5%</td> <td></td> </tr> <tr> <td>8</td> <td>Marital status</td> <td>15%</td> <td></td> </tr> </tbody> </table> | <u>Item No.</u>  | <u>Characteristic</u> | <u>Beginning</u> | <u>Changed to</u> | 2 | Relationship to head | 10% |  | 3 | Sex | 5% | 10% | 4 | Color or race | 15% |  | 7D | Decade of birth | 15% |  | 7Y | Year of birth | 5% |  | 8 | Marital status | 15% |  |
| <u>Item No.</u> | <u>Characteristic</u>   | <u>Beginning</u> | <u>Changed to</u>     |                  |                   |   |                      |     |  |   |     |    |     |   |               |     |  |    |                 |     |  |    |               |    |  |   |                |     |  |
| 2               | Relationship to head  | 10%              |                       |                  |                   |   |                      |     |  |   |     |    |     |   |               |     |  |    |                 |     |  |    |               |    |  |   |                |     |  |
| 3               | Sex   | 5%               | 10%                   |                  |                   |   |                      |     |  |   |     |    |     |   |               |     |  |    |                 |     |  |    |               |    |  |   |                |     |  |
| 4               | Color or race   | 15%              |                       |                  |                   |   |                      |     |  |   |     |    |     |   |               |     |  |    |                 |     |  |    |               |    |  |   |                |     |  |
| 7D              | Decade of birth   | 15%              |                       |                  |                   |   |                      |     |  |   |     |    |     |   |               |     |  |    |                 |     |  |    |               |    |  |   |                |     |  |
| 7Y              | Year of birth   | 5%               |                       |                  |                   |   |                      |     |  |   |     |    |     |   |               |     |  |    |                 |     |  |    |               |    |  |   |                |     |  |
| 8               | Marital status  | 15%              |                       |                  |                   |   |                      |     |  |   |     |    |     |   |               |     |  |    |                 |     |  |    |               |    |  |   |                |     |  |
| D               | Item 4 (color or race) was marked "Other" for more than 1 percent of the persons. (Changed to 99 percent.)  |                  |                       |                  |                   |   |                      |     |  |   |     |    |     |   |               |     |  |    |                 |     |  |    |               |    |  |   |                |     |  |
| E               | The blank rate for the housing item listed below exceeded the allowed tolerance:  |                  |                       |                  |                   |   |                      |     |  |   |     |    |     |   |               |     |  |    |                 |     |  |    |               |    |  |   |                |     |  |

<u>Item No.</u>	<u>Characteristic</u>	<u>Beginning</u>	<u>Changed to</u>
A	Number of units at address	15%	
H1	Telephone	25%	50%
H2	Access	10%	50%
H3	Kitchen facilities	10%	50%
H4	Number of rooms	10%	50%
H5	Water supply	10%	50%
H6	Flush toilet	10%	50%
H7	Bathtub or shower	10%	50%
H8	Basement	15%	50%
H9	Tenure	10%	50%

<u>Flag</u>	<u>Condition</u>			
<u>E--</u> <u>Cont.</u>	<u>Item No.</u>	<u>Characteristic</u>	<u>Beginning</u>	<u>Changed to</u>
	H10a	One-family home	15%	50%
	H10b	Commercial unit on property	25%	50%
	H11	Value	25%	50%
	H12	Contract rent	25%	20%
	B	Type of quarters	25%	
	C	Vacancy status	25%	50%
	D	Months vacant	25%	40%

- G More than 25 questionnaires had block numbers with less than three digits. (NOTE: Talled only for "blocked" ED's.)
- H More than 25 questionnaires had serial numbers with less than three digits.
- I There were more than 25 questionnaires for which either page 2 or page 3, or both pages 2 and 3, were not read on FOSDIC.
- K There were more than 60 indexmarks missed.

In general the diary flags were reviewed in alphabetic order. Flags I and K, however, indicated that the microfilm was not processed correctly and therefore that any information shown on the diary about the ED was suspect. In these cases the ED was remicrofilmed, but it was found useful to review flags C, D, and E so that individual questionnaires could be corrected if necessary before the ED was remicrofilmed.

If an ED required remicrofilming after review and correction either by a clerk or a subject-matter analyst, this was requested. Even though the problems in the original work unit were corrected and properly processed through FOSDIC, it still was possible for the ED's in the remicrofilmed work unit to have error flags. In these cases the latest diaries were compared with previous diaries. Unless the blank rates for the problem items had increased, the questionnaires were not reviewed again and the ED was released for sample coding. Because of the many possible ways that an ED could be handled, some were remicrofilmed many times; many ED's which had been accepted and released later were determined to require remicrofilming. This was caused by released ED's being missent to be remicrofilmed, or by a number of other factors which might cause error flags to appear during FOSDIC reprocessing.

Some released ED's had to be recalled for administrative splitting. Once an ED had been released for sample coding, it was difficult to retrieve both the short questionnaires which had been sent to storage and the long questionnaires in the coding operation in order to reconstruct the ED.

Occasionally two or more ED's had been microfilmed as one. When this was discovered, the questionnaires for the ED's were sorted, placed in separate cartons, and remicrofilmed. When only part of an ED had been microfilmed, the missing questionnaires were sought in the ED

cartons of questionnaires that had been microfilmed before and after the partial ED. If found, the partial ED was completed, and both it and the ED(s) which originally contained the missing questionnaires were sent for remicrofilming and FOSDIC processing. If the missing questionnaires could not be found, the partial ED was held in a "lost and found" unit until it was properly allocated or the State was closed out. (See below.)

### 100-Percent Diary Closeout

Top priority in processing the results of the 1970 census was given to the final population counts for each State required for congressional apportionment. These counts had to be completed within 8 months of the census date (April 1) and transmitted to the President. In order to meet this deadline a closeout date was set for each State, so that at that point 100-percent data corrections would be halted.

Generally, when the number of unacceptable ED's for a State was reduced to 100, or its closeout date was imminent, all the clerical processing for that State was transferred to a special closeout unit where the following steps were taken: A disposition list of the unaccepted ED's was prepared, indicating their status with regard to microfilming, FOSDIC processing, and subsequent review by clerks or subject-matter analysts. New tolerances allowed during closeout were applied to the diaries for these ED's, and any ED's passing the adjusted criteria were released. Within the allotted time the remaining ED's were repaired as well as possible, questionnaires were retrieved from the "lost and found" unit, and the ED's were remicrofilmed. When a time deadline was reached, any ED's still on the disposition list were checked against the diary control list--to make certain that no duplication had occurred--and against an extract of the Bureau's geographic reference tape--usually by telephone between Jeffersonville and Suitland. (The geographic reference tape extract included population and housing counts that were shown on the breaker sheets as well as the counts arrived at in FOSDIC processing.)

The first States (Montana, Vermont, and Wyoming) were closed out by regular diary-review procedures on August 1, 1970, at Jeffersonville. Shortly thereafter, the closeout operation was instituted with a staff of two clerks. At the peak of 100-percent closeout approximately 45 clerks and three supervisors were employed, and from September through mid-December 1970 closeout personnel worked 10 hours a day, 6 days a week.

In mid-October emergency procedures were instituted so that final population counts for all States could be completed in November. Lists of ED's having a difference of 500 persons or more between the count shown by the computer and that shown on the breaker sheet were prepared clerically in Suitland, and clerks in Jeffersonville compared these counts with the ones found on the address registers for the ED's in question to ascertain which totals were more nearly correct. At the same time population totals were determined for ED's listed on the geographic reference tape but not appearing on diary control records as having been processed through the

FOSDIC operation. The final population counts for 10 States were obtained by these emergency procedures, and the final counts for five other States were obtained by a combination of regular and emergency procedures. At no time did these emergency procedures apply to more than a small percentage of the ED's for a State. Once these counts had been obtained, however, closeout efforts continued, but they were concentrated on making internal corrections not affecting the State totals and on microfilming ED's as necessary. (Remicrofilming of ED's for 100-percent data ended in mid-December.)

The greatest number of problems were caused by ED's that were split in the field or in processing, either because they presented geographic problems unknown at the time they were delineated or because they proved to contain more population or housing units than could be processed on the computer. The chances for error in processing these split ED's were numerous, especially given the time constraints of the census. Each portion of a split ED required special handling with regard to geographic identification, breaker sheet preparation, and microfilming; and additional handling for FOSDIC and computer processing. Some splits were not reflected in the Bureau's master reference file and therefore had no geographic codes whereby they could be matched to the geographic reference tape. Many of the large discrepancies between the computer population counts and the counts on the breaker sheets for an ED occurred because the ED had been split, erroneously duplicated, or lost in preliminary processing.

#### Field Reinterview

An examination of field records indicated potential problems with the number of persons enumerated on sample questionnaires in certain ED's. As soon as 100-percent diary review of these ED's was completed, the address register for each was obtained and the sample count was verified. If the 20-percent sample population count still differed significantly from the established tolerances, the ED was resampled. By constructing a new address register, one-fifth of the households enumerated on the short questionnaires were designated as sample households, and four-fifths of those in the original sample were designated as 100-percent households.

Sample questionnaires were prepared and sent to the field for reenumeration, with instructions to enumerate the household as it was found, not as it existed on April 1. Upon completion, the questionnaires were returned to Jeffersonville, where the new sample questionnaires were substituted for the old ones (which then were destroyed). The 100-percent data from four-fifths of the old sample questionnaires that were designated as 100-percent households were transcribed to short questionnaires. All of the new questionnaires were assembled in new ED cartons and forwarded to the next step in processing--separation of the questionnaires preparatory to sample microfilming. A staff of as many as 15 clerks handled 849 ED's in this manner in the fall of 1970.

#### Block Statistics Diary Review

The purpose of the block statistics diary review was to ensure that 100-percent population and housing data were tabulated to the correct census blocks for all urbanized areas and contract block cities.

During the second-count computer processing, which produced 100-percent tabulations for census tracts, the computer compared the address serial number (for areas covered by address coding guides) and, for all areas, the block and the ED numbers attached to the data for each housing unit in blocked ED's (i.e., ED's in which the questionnaires were coded to census block) with the geographic reference tape (GRT). If, according to the GRT, the address serial number was unacceptable for the given block or the block number was wrong for that particular ED, the computer transferred that housing unit's data to an acceptable block. So that the changes could be reviewed clerically, the computer produced a diary of its actions. The diary summarized the data for the ED, listing the blocks in which the changes were made; in addition, it gave the address serial number for each household affected and the percentage of housing units and/or group quarters in which block changes, imputations, or allocations of data had been made, indicating the nature of these by "flags."

A clerical review unit was established in October 1970, with seven clerks and a supervisor. During classroom training sessions the staff was taught how to read maps, interpret boundaries, use address registers, and process the block diary review listings. Because an insufficient number of listings were received between November 10 and December 23, 1970, part of the clerical staff was shifted to another operation. This necessitated some retraining when the workload increased. Tight scheduling, late receipt of the listings, and delays in obtaining reference materials in use in other processing operations forced the clerical staff to work overtime to meet the schedule. The staff reached a peak of 65 persons in February 1971, and the operation was completed in July 1971.

In reviewing the listings, the clerks checked the affected address serial numbers against the address registers, and the original questionnaires where necessary, and compared the addresses with maps to verify the block numbers. If the new block number assigned by the computer was wrong and the correct block number existed within the ED, the clerks ordered the data transferred to the correct block number. At this stage of processing it was no longer possible to move the data from one ED to another, even if the block number was correct and the ED number was not. If the computer assignments within the ED became heavily concentrated and distorted reality, the housing units were distributed to other blocks in the ED. If any of the households in question also were in the sample, steps were taken to ensure that their sample data were coded and tabulated for the same ED and block as the 100-percent data had been.

Special attention was given to population and housing units that had no block numbers. The data for these had been distributed by the computer among the blocks in the ED. If the ED had only one or two legitimate blocks, it sometimes was necessary to accept additional blocks in the ED (as long as the blocks were in the same tract) to assure a reasonable distribution of the data.

Each clerk's work was verified 100-percent until it was determined to be of acceptable quality, and verification then was reduced to a sample. The corrections made in the clerical review were edited by computer to detect any possible errors.

A total of 148,331 blocked ED's were listed on the block statistics diary printouts. Of these, 90,834 ED's required clerical review, and 21,634 corrections were made within the limitations prescribed by the program. These were transcribed to punchcards for computer use.

### Sample Diary Review

Upon completion of the 100-percent diary review in December 1970, the staff which performed that operation was trained for sample diary review. As many clerks as possible--particularly those with sample coding experience (see p. 15 for a description of sample coding which preceded the diary operation)--were borrowed from other processing operations, but the total size of the staff was restricted by space limitations until June 1971 when additional space and manpower were made available. Except for a short period in February 1971, when sample processing was halted because of FOSDIC problems, sample diary review was carried on between February and September 1971.

As sample diaries and FIT's were received (in duplicate) for sample diary review, they were recorded by ED in control books which contained listings of all the ED's to be accounted for. Each movement of an ED through the sample diary review processing, including remicrofilming, was noted in these records.

Screening.--The diary and FIT for each computer processing unit (the sample data for up to 600 ED's) were screened, ED by ED. Original or remicrofilmed ED's without flags or with one or more flags within specified tolerance levels (see below) were accepted, and their cartons of questionnaires were released for storage. Overall, 70 percent of all the ED's screened (310,881, including remicrofilmed ED's) were accepted. The diaries and FIT's for the remaining ED's were examined to determine (a) what type of review was needed and (b) what further action was necessary before review began (for example, locating and annotating the original diaries for ED's that had been remicrofilmed; obtaining address registers; checking the diaries of adjacent ED's when it appeared that data for one ED had been merged with those for other ED's; and, when needed, ordering the remicrofilming of the questionnaires for an ED).

The materials needed by the diary review clerks, as indicated by screening, were gathered in a staging area. Diaries were inserted in the appropriate ED cartons; all the cartons for one State were sorted into five categories, according to the flags appearing on the diaries. A portion of the staging area was devoted to a "lost and found" operation in which stray breaker sheets and questionnaires, incomplete ED's, etc., were accumulated and redirected. After processing was completed for a given State, any remaining "lost and found" material for that State was filed. This screening and staging of the diary materials into complete units before they were given to the review clerks virtually eliminated the slow turnaround of work units which had been a problem during the early stages of the 100-percent diary review. During staging it also was possible to hold certain units while their tapes were being rerun on the computer, in anticipation that more ED's could be accepted after reprocessing and thus reduce the remaining diary review workload.

Review.--Clerks reviewed the diaries and accompanying materials in State order according to the respective deadlines for completing individual States. Where possible, they corrected the conditions which caused the appearance of flags on the diaries, and referred technical problems either to subject-matter specialists or systems analysts. (Some clerks were given the specific task of darkening the faintly coded FOSDIC circles which were the cause of many diary flags.) After correction a decision was made whether the ED could be accepted, or whether it would have to be remicrofilmed and reprocessed through FOSDIC. Some of the principal review actions were the following:

1. A and B flags.--The census sample was so designed that every fifth household was to be enumerated on a sample questionnaire; therefore there should have been sample data for 20 percent of the households in each ED. The appearance of an A or a B flag indicated, respectively, that the number of persons or housing units in the sample varied from the sample breaker-sheet counts by more than 12 persons or 5 housing units, and that these figures varied by more than 5 percent from the sample count totals shown on the breaker sheet for the ED. In general, the clerk accepted these flags if the sample totals--depending on the size of the sample--were as follows: (a) population, 15 to 25 percent of the 100-percent population count, and (b) housing units, 15 to 25 percent of the 100-percent housing count for the ED as found in the sample control records. If these tolerances were not met, the ED was referred to a special group of clerks who performed the following actions:

a. If the ED was short of sample population or housing units, the questionnaires in the ED carton were matched with the corresponding address register to determine whether sample questionnaires were missing. If they could be located, the ED was reassembled and remicrofilmed; if the missing sample questionnaires were not found, the ED was stored with the expectation that as processing of the State continued the missing questionnaires would appear. Conversely, excess questionnaires in an ED carton were removed and stored, and the ED was remicrofilmed.

b. As the processing of a State neared completion, those ED's that were still incomplete were reviewed by a subject-matter analyst to determine final disposition.

2. M flag.--The appearance of an M flag on the diary indicated that seven or more items of sample population data had been allocated by computer for each of 10 or more persons in a particular group quarters (such as a college dormitory, rooming house, or institution) and that this number of persons constituted more than 25 percent of the occupants (whether inmates or staff). These situations were resolved by transcribing to the blank pages entries from completed questionnaires for the same type of group quarters.

3. Individual items.--In general, flags appeared on the diaries for the individual population or housing items on the sample questionnaires if (1) there were supposed to be more than 20 responses for the given

item (such as a person's age, veteran status, etc., or the number of bedrooms or complete bathrooms in a housing unit) for the entire ED, and (2) the number of responses that had to be allocated by computer exceeded a specified ratio. For example, if there were 20 or more housing units in the ED enumerated on 5-percent questionnaires, and data for clothes washing machines (item H27a) were allocated for more than 50 percent of these cases, a flag appeared.

4. Other situations.--Problems caused by missing or duplicate pages on questionnaires, substitute questionnaires, cases in which the FOSDIC operation did not recognize the existence of data for a person or a housing unit, or other aberrations which caused differences in the control counts were resolved by referral to analysts and/or by remicrofilming and reprocessing the ED's in question.

Reconciliation--Finally, the entries on the control listing for 100-percent and sample data were compared to make certain that both types of data had been obtained for each ED. This comparison revealed most remaining discrepancies for a State and allowed their resolution before final sample tabulation was begun.

### Sample Diary Closeout

State-by-State deadlines were established for sample diary review so that data could be processed by the computer in time to meet the Bureau's time schedule for publishing reports. As these deadlines were reached, closeout procedures were begun.

First, when the original microfilming of a State's sample questionnaires was nearing completion, the Suitland office was notified and arrangements were made for priority handling of the final rolls of microfilm through FOSDIC and the computer. When the resultant diaries were received in Jeffersonville, an "end of State review" was begun for that State. This consisted of identifying, locating, reviewing, and remicrofilming if necessary all remaining unaccepted ED's in that State. When the "end of State review" was completed and remicrofilming was accomplished, the Suitland office was notified to put the State into closeout status. This involved priority processing of all remaining remicrofilmed work for that State through FOSDIC and the computer and producing a disposition listing which identified ED's that (1) were listed on two or more diaries, (2) were identified geographically but had no data, (3) had data but no geographic identification, (4) did not meet the established tolerance levels, (5) had sample data but no 100-percent data, (6) contained data that were improperly read by the computer, or (7) had data imputed (supplied) that would have to be compared with other 100-percent and sample figures to make certain that the imputations were appropriate.

Upon receipt of the disposition listing in Jeffersonville all work on that State was transferred to a closeout unit where the listed ED's were reviewed and corrections made. Normally, closeout began when the number of unaccepted ED's was reduced to approximately 100, but for some of the larger States the number of unaccepted ED's was as large as 400.

In mid-May 1971, additional guidelines were created to expedite the closeout operation. It was decided that if, after a second microfilming, the computer still found the sample data for an ED out of tolerance in a State containing at least 1,000 ED's, that State could be closed out with as many as 10 ED's either out of tolerance or containing other problems.

The initial coding and microfilming were completed for so many States at the same time that there were 28 States in the initial closeout phase. As a result, in early June 1971 it was decided that no additional States would be put into closeout status until only 10 of the 23 States in closeout at that time remained. After mid-June the size of the work units transmitted to FOSDIC and the computer was decreased in order to lessen the turnaround time between microfilming and the receipt of diaries in Jeffersonville. In August an additional 5-percent tolerance was granted for all items in the sample diary closeout operation. In late August it was decided to remicrofilm all ED's reported as "never received" at the computer. This microfilm was held in Jeffersonville and sent to Suitland for use as needed.

Closeout of "edit drops" (ED's containing data that had not been read by the computer) and correction of faulty FOSDIC runs often were delayed because of FOSDIC and computer work backlogs, but this was changed in early September when the sample diary operation received first priority on the computer. In mid-August a closeout schedule was established for each week. This schedule called for final closeout during the week of September 14, 1971. With the exception of one ED each in New York and Illinois, this schedule was met. (These two States were closed out on September 20, 1971.) The closeout staff continued to investigate problems involving population and housing counts until the unit was disbanded in December 1971.

## SAMPLE CODING

### Introduction

After the pages containing 100-percent data on all of the household questionnaires had been microfilmed, the sample questionnaires were separated from the others and reassembled by ED in State and county order to have the responses to certain sample items clerically coded in FOSDIC-readable form before the sample questionnaires were microfilmed and their data processed. This clerical coding operation, which is described in the following sections, fell into three categories: General, which covered a variety of subjects such as place of birth, income, and rent and utility costs; place of work (POW); and industry and occupation (I&O). Each of these coding procedures was subjected to verification and other quality control measures (including at times the comparison of three separate codings of the same data where necessary) and, when the quality was unacceptable, rectification of coding errors.

The total operation involved approximately 250,000 ED's, and data for approximately 40 million persons and 13.7 million housing units. Sample coding began in May 1970 and was essentially completed in June 1971.

**Coding Procedures**

General coding.--The purpose of the general coding operation was to examine and code the responses to the following items on the sample questionnaires, and edit and correct the respondents' or enumerators' entries if necessary:

Item	Description and coding action
a5	Address serial number (edit or enter 3 digits)
B	Type of unit (housing unit or group quarters) or continuation of another questionnaire (1 digit)
2	Relationship to head of household, family relationship, or type of group quarters (2 digits)
4	Other race, or Indian tribe (2 digits)
Person	Person number
13	Place of birth (3 digits)
14,15	Parents' birthplace (3 digits)
17	Mother tongue (2 digits)
19,36	Place of residence 5 years ago (5 digits or 2 digits)
24	Marital history (3 digits)
27	Vocational training (edit only, 1 digit)
40,41	Earnings and income (18 digits)

H12 Contract rent (3 digits)  
 H13 Cost of utilities not included in rent (10 digits)

(For detailed information on the individual items, and how each was coded and tabulated, see chapter 15.)

In addition, a separate unit within the general coding operation was responsible for coding questionnaires from the States of Arizona, California, Colorado, New Mexico, and Texas as to whether or not the persons listed had Spanish surnames (a 1-digit code). The Bureau prepared its own reference materials and coding guides for all of these operations. The final step in general coding was the placing of a 2" x 1/2" strip of black tape on each questionnaire to mark the last page of a set containing coding; in the microfilming operation, which followed, this was a signal to the camera mechanism to move to the next questionnaire.

Place-of-work coding.--The 1970 census was the first in which streets and building locations were geographically coded. Figure C shows the place-of-work inquiry, item 29c, together with its coding boxes, as they appeared on the 15-percent sample questionnaire when the response was completed and coded.

**Figure C. Place-of-Work Items and Coding Boxes on 15-Percent Sample Questionnaire**

QUESTIONS 29 THROUGH 41 ARE FOR ALL PERSONS BORN BEFORE APRIL 1956 INCLUDING HOUSEWIVES, STUDENTS, OR DISABLED PERSONS AS WELL AS PART-TIME OR FULL-TIME WORKERS

**29a. Did this person work at any time last week?**

Yes--Fill this circle if this person did full- or part-time work (Count part-time work such as a Saturday job, delivering papers, or helping without pay in a family business or farm, and active duty in the Armed Forces.)

No--Fill this circle if this person did not work, or did only own housework, school work, or volunteer work.

↓ Skip to 30

---

**b. How many hours did he work last week (at all jobs)?**  
 Subtract any time off and add overtime or extra hours worked.

1 to 14 hours       40 hours  
 15 to 29 hours       41 to 48 hours  
 30 to 34 hours       49 to 59 hours  
 35 to 39 hours       60 hours or more

---

**c. Where did he work last week?**  
 If he worked in more than one place, print where he worked most last week.  
 If he travels about in his work or if the place does not have a numbered address, see instruction sheet.

(1) Address (Number and street name) 1401 Some St

(2) Name of city, town, village, etc. Amystown

(3) Inside the limits of this city, town, village, etc.?  
 Yes  
 No

(4) County River

(5) State Pa.      (6) ZIP Code 19000

**29c.**

A	○	○	○	○
0	○	○	●	○
1	○	○	○	○
2	○	○	○	○
3	○	○	○	○
4	○	○	○	○
5	○	○	○	○
6	○	○	○	○
7	○	○	○	○
8	○	○	○	○
9	○	○	○	○

} ZIP code

---

**29c.**

0	○	○	○	○
1	○	○	○	○
2	○	○	○	○
3	○	○	○	○
4	○	○	○	○
5	○	○	○	○
6	○	○	○	○
7	○	○	○	○
8	○	○	○	○
9	○	○	○	○

} Street code or universal area code (UAC)

---

**29c.**

0	○	○	○	○
1	○	○	○	○
2	○	○	○	○
3	○	○	○	○
4	○	○	○	○
5	○	○	○	○
6	○	○	○	○
7	○	○	○	○
8	○	○	○	○
9	○	○	○	○

} Structure number (street address) code

On entering the POW coding operation, ED's were sorted into two types--those inside the 243 SMSA's, covered by the Bureau's place-of-work coding guides, and all others--and (during the last months of coding) assigned to clerks specializing in handling one type or another. For ED's outside these SMSA's the clerk entered for the place of work five zeros in the ZIP code field and a 5-digit universal area code (UAC) which identified the State, county, central city or other place of 20,000 inhabitants or more in 1960 (or at a subsequent special census), or selected minor civil division (in New England). For the ED's within the SMSA's, if the person's place of work was inside the SMSA and inside the area covered by POW guides, the clerk entered a 5-digit ZIP code, a 5-digit street code, and a 6-digit structure number. If the place of work was inside the SMSA but outside the POW coding guide area, only ZIP and UAC codes were entered. If the person worked outside the SMSA in which he resided, the clerk entered five zeros for ZIP code and the UAC code for the place of work unless this place was in another SMSA which, together with the residence SMSA, constituted a "commuter shed." In this case the clerk entered a full 16-digit code. (A commuter shed is an area containing at least two contiguous SMSA's in which, in 1960, at least 7.5 percent of the work force in one SMSA commuted from homes in an adjoining SMSA. When POW coding began, the commuter shed definition was applied to any SMSA which either received 7.5 percent of its workforce from one or more adjoining SMSA's or had at least 7.5 percent of its employed workers commuting to adjoining SMSA's to work. In March 1971 this definition was narrowed to include only the SMSA's receiving workers.)

The coding clerks were provided, as needed, with POW coding guides, telephone directories, national ZIP code directories, and lists for assigning UAC numbers for places of work shown as military installations, colleges or universities, or simply place names. If the information contained in the response to question 29c was insufficient to allow coding by reference to one or more directories, or if the entries were blank, the clerk referred to question 33a ("For whom did he work?") to see if an employer's name was listed there. If so, the clerk looked for this name in appropriate directories to secure a codable address. Special codes were assigned to designate place of work for persons in the Armed Forces with no geographic information given, for persons overseas or at sea, or for persons working in known foreign countries. The clerks followed a decision logic chart. Doubtful cases, or cases which could not be coded on the basis of information available to the clerks, were assigned to referral clerks for resolution. (In late January 1971 it was decided to allow an ED with three or less problem cases to pass without resolution; in May, in order to meet June coding deadlines, this maximum was increased to five.)

The POW coding operation was hampered initially by an insufficiency of reference materials, which meant that coding of some ED's was delayed. Several POW coding guides were 3 to 4 months late in arriving or were incomplete, and telephone directories were not in adequate supply until January 1971. As in the other coding operations, there were not enough rolling bins available, so that space was taken up with stacked materials. The 17,000 copies of POW coding guides and 11,000 telephone directories alone required a large storage area.

Industry and occupation coding.--The purpose of industry and occupation coding was to enter on the sample questionnaires numeric or alphabetic codes based upon the responses pertaining to all persons listed except those who were under 14 years of age or for whom the answer "1959 or earlier" or "never worked" had been given to question 32 ("when did he last work at all, even for a few days?") Figure D shows the I&O inquiries, items 33 (industry), 34 (occupation), and 38 (activity 5 years ago), together with their coding boxes, as they appeared on the 5-percent sample questionnaire when the responses were completed and coded. (Items 33 and 34, but not item 38, also appeared on the 15-percent sample questionnaire and were coded in the same manner.) (For details on the individual items, and how each one was coded and tabulated, see chapter 15.) The questionnaires to be coded for each ED within a given State usually were received from one of the other coding operations (general or POW) and were moved through the I&O clerical coding operations in three stages--precoding, coding, and postcoding. (See discussion of quality control below.)

The coders had two basic reference sources:

1. The company name list, generated from the Bureau's 1967 Economic Censuses and updated in November 1969. This list consisted of approximately 650,000 company names printed out on about 8,000 pages. It did not include categories not covered in the economic censuses, however, such as banks, professions (lawyers, doctors, etc.), schools, hospitals, and farms, and for some industries was limited to companies above certain employment levels. (When coding began there were only 100 to 150 copies of the company name list for each State. As reproduction facilities at Jeffersonville were limited, a number of copies were obtained from private contractors.) For firms and industries not found in the company name list, the coders went directly to--

2. The alphabetic index of industries and occupations. The coders looked up the industry reported on the questionnaire, compared the occupation reported with those listed in the index under that industry and, if a match occurred, coded the questionnaire with the code provided in the index.

If no match was possible, the questionnaire was referred to a subject-matter specialist for assignment of codes. A number of changes were made to the alphabetic index and to the company name list as new data became available from coding already completed. The index was reprinted in January 1971.

Clerks were also provided with other reference materials, such as a coding guide for institutions, which was issued to them in parts being prepared in the order in which the States were expected to be processed. As this State-order was changed a number of times as processing progressed, the coders sometimes did not have the guides they needed, and delays occurred. These reference materials also were amended during coding. In all, over 1,000 pages of changes were issued to clerks.



## Quality Control

In the 1960 censuses of population and housing, sample coding consisted of manually assigning FOSDIC-readable codes for the written answers to certain questions on the mailed-in questionnaires. A three-way independent verification system was developed to control the quality of the coding. The system consisted of independently coding a sample of the questionnaires by three clerks, matching the three sets of codes, and assigning errors to the clerk who disagreed with the other two clerks.

The same type of coding was planned for 1970: An independent system, similar in principle to that used for 1960, was used during the first 6 months of the 1970 coding operation to assure an acceptable outgoing error rate. One significant change was made in the 1970 procedures: There was still an independent coding by each of two clerks, the precoder and the coder, but only if these two clerks differed in the codes assigned were the disputed items independently coded by a third clerk, the postcoder. Prior to coding, each questionnaire in the quality control (QC) sample was photocopied, and this copy was used by the precoder and postcoder in determining the codes. Codes were entered on transcription cards and, since no clerks had knowledge of the codes assigned by others, the coding was independent.

In 1970 mechanization was introduced for matching the codes assigned by the three different sets of coders. The process was designed around the use of a mark-sense card which could carry both punched and pencil-marked data. Seven different formats of these mark-sense cards were used. (See figs. E and F for example.) The cards were read by a mark-sense reader (a modification of a standard punchcard reader) which was linked to a small computer in Jeffersonville.

The quality control operation was carried out as follows: A sample of questionnaires was selected from each ED. A 1-in-19 sampling rate was used initially for POW and I&O coding, and a 1-in-38 rate for general coding. In August 1970, after the coding production rates had increased, the sampling rate was changed to 1-in-39 for POW and I&O coding and 1-in-78 for general coding. Every questionnaire selected was verified for place-of-work and industry-and-occupation coding, and every other questionnaire selected was verified for general coding. The selected questionnaires were photocopied, and a clerk marked on the mark-sense sample selection card for each one the questionnaire identification, the persons or housing unit to be coded, and the type of coding required. These cards were fed into the mark-sense reader, and the identification data were transferred by computer punching and printing onto two sets of other mark-sense coding documents--one set for the precoder to use and the other for the transcriber.

A precoder, using the photocopy of the sample questionnaire, coded the responses on one set of the mark-sense cards. A transcriber copied onto the other set of transcription cards the codes entered on the actual questionnaire by the production coder. An identification card was added to each set of documents to identify the precoder and the coder. The two sets of documents then were compared on the computer.

The output of the computer comparison was (1) a printed listing of those items with code differences and (2) a third mark-sense coding card for each person or housing unit with at least one code difference; this card indicated for postcoding only those questions with a code difference. The postcoder, using the photocopy (uncoded) of the questionnaire, coded all questions in which differences were involved and entered his codes on the third set of mark-sense cards.

A three-way comparison of the codes then was made by the computer. Two codes in agreement were considered correct, and the clerk with the difference was charged with an error. In the case of a three-way difference, as in 1960, no one was charged with an error. The output of the three-way comparison was an error listing and, for each clerk, a summary card containing detailed error information. The clerks' summary cards then were transmitted to Suitland by datalink for input to a computer which made accept/reject decisions on the work and produced daily and cumulative records by clerk, coding unit, and coding type.

Numerous problems were encountered almost immediately in using this system. Although the staff had been trained and the computer had been installed, the first preprinted mark-sense cards to be used as input to the computer did not arrive until mid-June 1970, 3 weeks after coding had begun. To begin the operation, blank cards were used, with a two-part plastic template. Using the template was slower than using the preprinted cards because the cards had to be aligned manually under the template to permit precise marking. Also, a card had to be inserted in and removed from the template for each sample-selected questionnaire. Even a slight variance in card placement caused misplaced marks and consequent error. The delays caused by having to use blank mark-sense cards created a backlog of QC work which was never overcome. A considerable time lag resulted between the coding and the QC decision, and feedback to the coders, precoders, and postcoders was almost nonexistent. The mechanical QC system was planned to produce a decision on a coder's work in 2 days, so that errors or misconceptions could be corrected promptly and poor coders could be retrained or replaced as soon as possible. Even when the mechanical system was working, no decisions were possible in less than a week. Therefore, in October 1970, a decision was made to adopt a different system for controlling coding quality. The second system, which is described below, went into effect in November 1970, after approximately 25 percent of the coding had been done.

Dependent verification, in which the verifier directly saw and could be influenced by the coder's work, was instituted in general and place-of-work coding. After production coding for an ED was completed, a verifier selected a sample of questionnaires. The sampling rate was 1-in-19 for general coding and 1-in-9 (reduced to 1-in-19 in February 1971 and 1-in-38 in May 1971) for POW coding. The verifiers checked the sampled questionnaires, corrected any errors they detected, and prepared a report of their findings for each ED. Reverification was performed on a 1-in-11 sample of the verification sample to control the verifiers' error rates. Differences between the coder's and the verifier's work were adjudicated by the



1971, a further check was instituted: A 5-percent sample of the postcoders' work was selected and independently postcoded in order to evaluate the postcoder's accuracy.

In all three types of coding, one copy of the verification report was sent immediately to the coder (and precoder in I&O) for feedback of errors. Another copy of the verification report was sent to a control unit where daily accept/reject decisions were made for each ED and records were kept for each coder. The decision to accept or reject the ED was based on the coder's work only. The decisions and records were posted manually until March 1971, when a computer system was initiated. This system was designed to maintain records similar to those produced for the clerk summary cards in the mechanical QC system.

Initially, all coders were considered qualified to perform the coding operations. If, in quality control, three ED's were rejected in any series of 10 ED's coded by the same clerk, that clerk was disqualified and either was given additional training or was released. Beginning in December 1970, however, a coder was disqualified as soon as he accumulated three reject decisions, regardless of the number of ED's reviewed. The decision tables used during quality control inspection were designed to allow a maximum error rate of 5 percent in general, 15 percent in POW, and 13 percent in I&O coding.

When the control records indicated that decisions had been made for an ED in all three types of coding, the ED was released either for microfilming or for rectification (100-percent reverification). Search teams located the appropriate ED cartons in one of several storage areas and delivered them to their proper destinations.

The quality of the sample coding improved steadily after the operations were stabilized and clerks began receiving regular feedback on their work. Between November 1970 and the end of coding in mid-1971, the average outgoing error rate for general coding moved downward from 1.4 percent to 1.0 percent; in POW coding, from 15.6 percent to 3.5 percent; and in I&O coding, from 13.2 percent to 4.8 percent.

## Organizing and Staffing the Coding Operations

Introduction.--During the planning phase two different organizational structures, both of which were eventually implemented, were considered. The first structure, which was established in May 1970, consisted of three consolidated coding branches and one processing branch. Under this structure the questionnaires for a selected group of States would be coded through general, POW, and I&O coding within a single branch. In addition, problems that arose would be resolved in that branch. This structure had several advantages: Movement of materials would be reduced (all coding for a State would be done in one room); personnel reductions would have a less severe impact on any type of coding operation; and, as a result of competition among branches, an esprit de corps possibly might develop and result in increased production. The alternative structure consisted of specialized branches--a separate branch for each of the three types of coding, as was used in 1960. The specialized-branch structure was implemented in November 1970 and was

used for the remainder of the operation. Among its advantages were that (1) changes were more consistent within a specific type of coding and (2) supervisors became more technically competent when not required to oversee a variety of procedures.

Staff requirements.--The 1960 census cost and production records for general and I&O coding indicated that in 1960-1961 production increased continuously for about the first 20 weeks of coding, during which time approximately 35 percent of the workload had been coded. At this point each clerk was coding about 875 persons per day. By the time 50 percent of the 1960 coding was completed, individual coder production reached a peak of about 920 persons per day. For the 1970 coding operation it was assumed that maximum weekly production would be about 80 percent (737 persons) of the 1960 rate. (Rates for POW coding were based on the general coding rates.)

It was estimated therefore that for 1970--regardless of how the coders were grouped to perform their duties--the following staffs would be required:

General coding would require a peak staff of 385 clerks in July 1970, maintain that level through March 1971, lower to a staff of 12 through April, and phase out in early May 1971. POW coding would require a staff of 140 in June 1970, reached a peak of 385 in July, and remain at that level through March 1971. I&O coding would require a staff of 46 in June 1970, 265 in July, and a peak number of 443 in August 1970.) It would remain at 443 through April 1971, and phase out in May. These estimated staff requirements included precoders, postcoders, rectifiers, transcribers, verifiers, and referral clerks (all described below), but the estimates did not include supervisory personnel.

It soon developed, for reasons described below, that sample coding would have to be carried on at a much lower level of staffing (and production) until the last part of 1970. The following table reflects actual staffing during the first half of 1971, and is representative of sample coding operations under full production conditions.

Type of coding	Average staff per week	Range, Jan.-May 1971 (excludes final week of coding)
General	362	476 (Apr.) - 253 (end of May)
POW	311	374 (Mar.) - 182 (end of May)
I&O	350	447 (Feb.) - 243 (end of May)

## Training

Training was staggered so that general coding production could begin immediately after the clerks' classwork was completed, so that a backlog of questionnaires would be available for POW coding which began a week later, and then for I&O coding the week after that. As each class of about 30 clerks was formed into a production coding unit, another class began which led to the formation of an additional production unit. Nine coding classes were conducted simultaneously during the peak training months of June and July 1970.

Representatives from Bureau headquarters conducted 1-week training classes for supervisors in May and early June 1970. Each supervisor also attended an 8-hour

management class taught by Jeffersonville personnel. Some of the supervisors, as well as several of the Suitland personnel, then acted as trainers for the coders' classes which followed.

Training of clerks began the week of May 25, 1970, but was suspended between late July and late September of that year; the last training classes were held in April 1971. The total number of classes and attendees (including 24 supervisors) were as follows: 49 general coding classes for 1,457 persons; 36 POW coding classes for 1,102 persons; and 52 I&O coding classes for 1,510 persons. Most of these classes lasted 4-1/2 days, but as coding progressed and additional instructions were issued--particularly for general and POW coding--the length of class time was increased by 1 or 2 days. In addition to the training classes, 2- to 3-hour refresher classes were held for some clerks in November and December 1970 to review instructions and revisions to procedures. In the latter stages of the coding operation (in 1971) direct on-the-job training was given to 22 place-of-work coders with satisfactory results.

The original training plans anticipated initial training for 1,213 coders plus an additional 15 percent for replacements. By the time coding training had been completed, 4,091 coders had received initial training, and 1,178 coders had received classroom refresher training. Approximately 14,000 man-days in excess of original estimates were required for initial coding training, and an additional 351 man-days were needed for refresher training. The volume and scope of the training required the services of a full-time coordinator and two or three clerks; final preparation of examination questionnaires and other training materials, such as the collation and assembly of approximately 2.6 million pages of instructions, exercises, etc., required the occasional services of 10 or more clerks. Generally, these clerks were new employees awaiting training, but frequently they were borrowed from production units.

For general and POW coding, prospective coders usually were allowed to repeat the course once if they failed to pass examinations; if they failed a second time they were assigned to noncoding work or released. For I&O coding, prospective coders failing one course were not given the opportunity to repeat it, but occasionally they were allowed to take general or POW coding classes for 1 further week in lieu of a noncoding job assignment.

Upon completion of their initial training as coders, those clerks who were selected for specialized work such as precoding, postcoding, referral, or correction were given on-the-job training as needed. The persons selected to act as precoders and transcribers were chosen randomly from each training class or coding unit. The postcoders were selected at random from trained coders who were willing to work evening shift hours. The referral clerks were selected from those coders (1) having the best classroom records or (2) having been reviewed most favorably early in production coding. The rectifiers were chosen from those coders having the lowest error rates.

### Organizing the Coding Branches

Between late May and early November 1970, there were three consolidated coding branches, with only slight

differences among their structures. Each branch contained one or more units of 20 to 40 clerks and one transcriber for each type of coding--general, POW, and I&O. Selected support functions, such as precoding and postcoding, were assigned to one of the branches. An I&O problem referral unit was established in one of the branches. This unit consisted of eight teams of problem solvers and clerks, some of whom had experience in the Bureau's Current Population Surveys; the teams worked in all the branches as necessary, solving problems. During peak operations there were approximately 70 persons engaged in I&O referral alone. In addition, each branch had its own general and place-of-work problem-referral unit, consisting of about 20 clerks each. Each branch was assigned certain States for coding and was provided with reference materials for those States.

The three coding branches were located in the same building, but in three different rooms, which were crowded at times of peak staff size. Because of the space problem and the limited number of reference materials, it was necessary to arrange the coders' desks in double rows with the coders facing each other. This arrangement facilitated the sharing of materials, but it also led to lost production time. As more space became available the desks were placed in single rows, facing in the same direction.

A logistical and quality control problem, which was never resolved entirely, was created because the precoders (of which there were 80 at the peak of operations), the photocopying equipment, and the computer-generated mark-sense cards which the precoders used were located for a time in three separate buildings. For a time, also, the precoders and postcoders were separated physically from the coders in order to avoid bias in the independent quality control program.

During the summer of 1970, the sample coding work force was held to a minimum of 600 persons. One of the coding branches suspended operations in August, except for several units engaged in a test operation. The daily flow of ED's released from diary review throughout June and early July was insufficient for the number of coders on duty. So that the remaining work force could be fully utilized, it was decided not to hold ED's until those for any given county or State were complete before assigning them to coders. It was necessary to code ED's whenever they were available until the flow of ED's from diary review increased to the point where work could be backlogged. The demands for an already limited supply of reference materials--especially guides and telephone directories for place-of-work coding--were difficult to meet under these conditions.

The situation was further complicated by the presence of ED cartons which often contained the sample questionnaires for as many as 10 to 20 different ED's. (Early in the processing operations, procedures were changed so that all of the questionnaires for one ED were placed in one carton together, but by that time a number of multi-ED cartons were already being processed.) Major problems had developed in the mechanical quality control system (see p. 19), and the coders were not receiving the feedback on their work needed for prompt adjustment of coding practices and prompt correction of miscoded questionnaires.

There was an overall loss of 62 percent of the coding staff when students working for the summer returned to school. Beginning in late September recruiting was begun to bring the coding staff up to 1,100 clerks by mid-October. (This figure was further augmented by the addition of extra clerks for I&O coding. The learning period for I&O coding was longer than for the other types and the independent verification system required additional coding; therefore it was necessary to offset the consequent lower production with more manpower.) By the latter part of October the 100-percent diary review was practically completed, and supervisory staff and bins of ED's once more were available for the coding operation.

In November 1970 the three consolidated coding branches were reorganized into three new branches, each one specializing in one of the types of coding, plus a coding analysis (quality control) unit and a referral unit (see p. 17) which handled the various problems that had to be resolved by subject-matter specialists. This arrangement was continued throughout the remainder of the coding operation.

Until February 1971 the coding was performed in the following order: (1) General, (2) POW, and (3) I&O. After that date the order was changed to (1) POW, (2) general, and (3) I&O. This change was made because the POW coding could most efficiently be done by coders who would concentrate on a single SMSA. This required that the POW coding operation be able to select the ED's to be coded as they were released from 100-percent and block statistics diary review. Based on production standards (see below) it also was expected that POW coding, being faster, could supply a greater backlog of work for general coding than general coding could for POW.

## Production

Standards.--The first benchmark standards against which coding production could be measured were based on the Bureau's experience in the 1960 census. These standards, used when 1970 coding began, were (per 8-hour man-day) as follows: 1,000 persons coded in general and POW coding and 904 persons in I&O coding. Data from production figures in the summer of 1970 indicated that the estimated production rates for general coding had been too optimistic. With the exception of the first 2 weeks of coding, when production was higher than expected, in no week was the estimated production rate met or exceeded by a substantial number of coders. In POW coding the production rate also was higher than expected during the first 2 weeks of coding. After this period the production rate leveled off, but the estimated rate generally was achieved.

In August 1970, eight management analysts and 20 clerks were formed into two teams which observed production. In November 1970, as a result of their observations, the benchmark standards per 8-hour man-day were revised as follows: In general coding, 800 persons (outside the five Southwestern States where Spanish surname also was coded, but for which a standard was not established); in POW coding, 800 persons for ED's inside SMSA's covered by address coding guides (ACG's) and 2,285 persons elsewhere; and in I&O coding, 706 persons.

Engineered standards, upon which incentive awards were based (see below), were established in January 1971. These were as follows:

	Number of persons to be coded per 8-hour man-day		
	General	POW	I&O
ED's inside SMSA's covered by ACG's. . . . .	774	842	873
All other ED's. . . . .	1,091	2,087	1,000

For the five Southwestern States where Spanish surname was coded, the standard for general coding was 593 persons inside the ACG SMSA's and 727 persons elsewhere. Standards also were established for verifiers and rectifiers. In February, as coding of the more populous areas began and the complexity of coding increased, the standards for general coding outside the SMSA's were lowered to 649 in the five Southwestern States and 842 elsewhere in the country. In March a standard of 739 persons per man-day was added for POW coding of ED's in nine large SMSA's where it frequently was necessary to refer to guides for two or more SMSA's in order to complete coding.

Incentive awards.--A system for providing incentive awards based on quality as well as production was implemented in January 1971, and supervisors, coders, verifiers, and quality control clerks were eligible. Re-verifiers were not included in the incentive plan but received appropriate pay adjustments based on experience and performance. Awards were given to clerks with production rates of 85 percent of the standard or more, provided their error rates stayed below a prescribed level (2.0 percent for general coding, 7.0 percent for POW, and 9.9 percent for I&O coding). Supervisors who spent at least 51 percent of their time in supervising coding were given incentive pay based on the performance of their units:

1. If the collective performance of the unit was 100 percent of the standard or higher, and at least 80 percent of the coders were qualified. (A coder was "qualified" if his production was a minimum of 70 percent of the standard and his coding error rate was no higher than the maximum established for incentive pay.)

OR

2. If a unit operating at less than 100-percent production raised its production level by at least 10 percent during a specified period.

At the same time warning letters were sent to clerks who failed to meet minimum standards for production and quality, or whose work fell below the level required for the incentive pay they were receiving. Continued disqualification led to retraining or replacement.

After production standards were imposed in January 1971, the POW coding operation exceeded 100 percent of the standard rate from March through late May 1971--except for 1 week at 96 percent--for ED's outside ACG SMSA's. For ED's inside the ACG SMSA's, POW coders exceeded 100 percent of the standard and continued to

exceed this figure--except for 1 week at 99 percent--through the end of the coding operation.

The following table reflects the production in the three types of coding and the manpower required:

Type of coding	Number of persons coded (rounded)	Number of coding man-days <sup>1</sup>	Coding completed during week of--
General	39,630,000	115,864	Jun. 7, 1971
POW	39,918,000	101,974	May 31, 1971
I&O	39,590,000	140,190	Jun. 14, 1971

<sup>1</sup>Includes verification and correction.

### Distribution

When the final coding, I&O, was completed for an ED, it passed to a special unit (created in November 1970) which acted as a distribution center. ED's were transmitted to and received from problem referral and rectification (see below), stored as needed between these steps, and transmitted to be microfilmed when complete. Clerks in this unit spent a large portion of their time determining the status of over 10,000 ED's for which quality control (QC) decisions were in doubt or had been lost because of changes in the QC system. If this status could not be determined, the ED was referred to a special dependent verification unit where all three types of coding on a sample of the questionnaires were checked and the necessary QC decisions made.

### Problem Referral

The referral staff included subject-matter specialists as well as clerks who used commercial directories, atlases, and other source materials in an effort to resolve coding problems. Within the referral unit as reorganized in November 1970, one section dealt with cases from general and POW coding; it contained an average of 30 clerks (60 clerks at the peak of operations) and processed 92,882 ED's. The other section specialized in I&O coding; this section contained an average of 60 clerks (128 clerks at the peak of operations) and processed 170,025 ED's.

When coding began all coding problems were referred, but in December 1970 an ED containing no more than two problem cases in any one of the three coding categories was allowed to bypass referral. In January 1971 the allowable number of cases was changed to three and remained that until the end of coding, except for POW coding where the maximum was increased to five in May 1971.

Ordinarily, all of the sample questionnaires in an ED were put through general and POW coding before moving to (or bypassing) referral. After solution of problems the ED was transferred to I&O coding. If further coding problems were encountered during the I&O phase, the ED was sent to the I&O referral unit after the I&O coding was completed. Since 83 percent of the ED's handled in I&O coding contained problems, a system of on-site referral was begun in February 1971, when 34 referral clerks were moved into the coding area to resolve problems while the

items were being coded (and before post-coding). This system was successful and substantially increased the work flow; it also provided fast feedback to the coders.

The referral operation was affected by the conditions which hampered sample coding as a whole--shortage of reference materials and lack of space. The space problem caused the referral clerks to be dispersed among a variety of locations, which complicated the movement of ED's and reference materials.

### Rectification

As the last step between I&O coding and/or problem referral and sample microfilming, certain ED's were channeled through the rectification unit. Its task was to correct according to current coding procedures the general, POW, and I&O coding in any ED which failed quality control for that particular type of coding, and to tally the coders' errors by type of error. Coders with low error rates were selected to be rectifiers. Beginning with a staff of less than 20 people in late October 1970, the number of rectifiers eventually reached peaks of 14 in general coding, 52 in POW coding, and 128 in I&O coding. Personnel changes in this unit were frequent. The staff was reduced as the quality of the production coding improved and work backlogs diminished. The unit rectified 7,885 ED's for general coding, 16,293 ED's for POW coding, and 28,019 ED's for I&O coding.

Between January and June 1971, clerks checked a 1-in-10 sample of the ED's in which POW and I&O codes had been corrected. This review, which was summarized weekly for management use, indicated that the rectification operation resulted in reducing the average outgoing error rate for rectified ED's from 10.7 percent to 6.4 percent for POW coding and from 6.4 percent to 2.3 percent for I&O coding.

## SUPPLEMENTAL FORMS AND COVERAGE IMPROVEMENT PROCEDURES

### Introduction

Several specialized or supplemental forms were used for the enumeration of persons not covered by the regular household census in the United States, Puerto Rico, and the outlying areas. These included form D-21, "Report for Military and Maritime Personnel"; form D-23, "Overseas Census Report"; and Form D-27, "Overseas Travelers' Report." A fourth form, D-20, "Individual Census Report," was used in particular situations where it was impractical to use a regular household questionnaire. Persons that might have been missed in the regular census operation were enumerated on several different forms: D-26, "Were You Counted?"; D-51, a questionnaire used in the post-enumeration Post Office check; D-910, "Record of Occupants," used in the National Vacancy Check; and a card entitled, "Please Make Sure I Am Counted in the Census." (For enumeration procedures, see chapters 5 and 7.)

All of these forms required special clerical processing. Most of this processing was done at the Bureau's Jeffersonville facility rather than in the district offices.

## Military and Maritime Personnel

Members of the Armed Forces living in barracks and other group quarters on military bases in the United States, Puerto Rico, and the outlying areas were enumerated on form D-21, "Report for Military and Maritime Personnel." The data from these forms were transcribed to regular FOSDIC-readable household questionnaires which were handled by the Bureau's district offices and processed in the normal manner. The persons enumerated were included in the preliminary population counts for the places where they were stationed.

The Department of Defense provided the Census Bureau with tabulations by age, sex, race, marital status, education, and State of "home of record" for members of the Army, Air Force, and Marine Corps stationed overseas as of April 1, 1970. These persons were not required to complete census questionnaires but were included in the apportionment counts for the States in which their homes of record were reported.

Crews of American flag vessels engaged in ocean-going, coastal, or Great Lakes commerce, and members of the Armed Forces and civilians stationed aboard U.S. Navy and Coast Guard ships in harbor or at sea on Census Day (April 1), also were enumerated on form D-21. When completed, these forms were collected and returned directly to Jeffersonville. On receipt, the forms were sorted by military or civilian vessel and checked off against control cards for the ships from which returns were expected.

Based on location reports, each nonmilitary vessel was classified as "U.S." if it was in a U.S. port or, if elsewhere, as "At Sea." (A vessel engaged in coastal trade was classified as "U.S." regardless of whether it was in port or not.) Each vessel classified as "U.S." was assigned a serial number, and the census forms for its personnel were coded to an enumeration district (ED) and a census block assigned to the pier area of the port where the vessel happened to be or, in the case of a coastwise vessel, its home port. The data on the forms were transcribed to regular short and long household questionnaires, following the prescribed census sampling pattern. A sample of the transcribers' work was verified to assure quality.

Based on reports from the Department of Defense, the Bureau assigned the crews of all military vessels not in the 6th or 7th Fleets of the U.S. Navy to their home ports regardless of their actual location on Census Day. The crews of vessels in the 6th and 7th Fleets were classified as part of the overseas population. (See below.) All personnel of the U.S. Navy aircraft mobile squadrons temporarily ashore on Census Day were assigned to the home ports of their ships. The data for all military and civilian personnel to be assigned to home ports were transcribed to FOSDIC-readable short and long household questionnaires and also coded to ED and census block in the same manner as the data for the crews of nonmilitary vessels. The questionnaires were placed in cartons and forwarded to a staging area where breaker sheets were completed and where the cartons were incorporated in the regular microfilming workload for the States to which they belonged.

The returns for ships "At Sea" then were inspected. Data for civilian personnel were transcribed to form D-23, "Overseas Census Report." The D-23 forms and the remaining D-21 forms for military personnel classified as "At Sea" were held for further processing. (See "Americans Living Abroad" below.)

## Overseas Travelers

Approximately 1,400 D-27 forms, "Overseas Travelers' Report," were received. The local addresses reported were coded to ED and the persons were searched for in the address registers. If not already enumerated, the data for the person(s) were added to the ED by listing them on a household questionnaire.

## Americans Living Abroad

Forms D-23, "Overseas Census Report," were completed for overseas land-based Navy and Coast Guard personnel, for military dependents, and for American civilians, including Federal employees and their dependents, living abroad for extended periods (as distinguished from tourists). If data were received for any of these persons on "Individual Census Reports," form D-20, they were transcribed to D-23 forms.

## Processing of Overseas and Military/Maritime Reports

Most forms D-23 were received from U.S. diplomatic and consular posts. Those forms for Federal employees and their dependents living abroad were sorted by country, and a count of the number of persons by "home State" was added to the State totals for apportionment purposes along with the counts of military abroad which were provided on tape.

The next step was to prepare figures for the published reports. There was a combined editing and coding of the D-21 and D-23 questionnaires, which started in the latter part of September 1970 and was completed in January 1971. All work was verified and the clerks were advised of their errors. Upon completion of the edit-code operation, the data were keypunched by work unit for input to the computer in Suitland, and the questionnaires were stored for the diary review operation.

Diary review clerks inspected the diaries resulting from computer processing, determined the acceptability of the data, and took corrective action where indicated. Corrections were verified and punched for further computer processing. Disposition listings, which reported the number of persons by country of residence, were reviewed in Suitland and compared with tallies of persons by home State. Only isolated problems arising from this review were researched in Jeffersonville.

A staff of 25 to 30 clerks in Jeffersonville processed the overseas and military/maritime questionnaires, which covered approximately 932,000 persons. In general, few difficulties were encountered in this operation once final procedures had been established.

## Late Receipts of Individual Census Reports and "Were You Counted?" Forms

Most of the ICR's (form D-20, "Individual Census Report") containing information for transients and persons who were away from home during the enumeration period were sent to the census district office of the person's permanent residence for processing. There the data were added to the appropriate household questionnaires if a search of the census records failed to find them already enumerated. A number of these ICR's were received in the district offices too late for processing, and were forwarded to Jeffersonville. The district offices also sent to Jeffersonville for processing the "Were You Counted?" forms (D-26), which consisted of the actual forms clipped from newspapers and, occasionally, typed lists and groups of questionnaires that had been employed instead of the form itself.

In June 1970, a staff of approximately 35 clerks was established in Jeffersonville to sort and process about 83,000 ICR's and 73,000 "Were You Counted?" forms which arrived there. (These receipts occasionally included late-arriving household questionnaires. Wherever possible these were inserted into the appropriate ED cartons, but if the ED was already being processed the questionnaire was treated as an ICR.)

The ICR's and "Were You Counted?" forms were grouped together for processing. Those forms with an ED number given were searched for in the address register and added to the specified ED if not found. All other forms with no ED numbers were processed if they represented 1 percent or more of the total population count of any place of 10,000 inhabitants or more, or the balance of the county.

For those areas having sufficient forms for processing, a sample of the forms was drawn for search against the census records. Upon completion of the search, the percentage of persons found was converted to a deletion rate and this rate was applied to the balance of the cases in the area. All cases not found in the sample search or not deleted by the application of the obtained rate were added to the largest ED's in the area. Additions to an ED were limited to 10 percent of the original population of the ED.

In the early stages of processing, questionnaires for these cases were added to the appropriate ED's and microfilmed. Later a revision in the procedure was made so that it would not be necessary to remicrofilm an ED in order to add these questionnaires. An imputation procedure for adding population had already been developed for the post-enumeration Post Office check (PEPOC) in the Southern States (see below), and that procedure was extended to this operation.

For each State outside the Southern region, a separate imputation deck was prepared. This deck consisted of a reel of microfilm containing the 100-percent data pages of all questionnaires in one large and one small ED randomly selected from that State. After the number of persons to be added to the area was determined, that number was divided by three to approximate the number of housing units these persons represented. The count of these "households" to be added to any given area then

was spread over the largest ED's in the area, not to exceed 10 percent of the population for any one ED. A breaker sheet was prepared for each of these ED's to indicate the number of households being added to the census. For each ED, the breaker sheet was microfilmed and households equal to the number shown on the breaker sheet were drawn in rotation from the imputation deck. The resultant data were added to and combined with the original data for these ED's.

Several problems were encountered in processing the forms. For example, several different kinds of "Were You Counted?" forms were received: In addition to the official Census Bureau ones, there also were forms clipped from newspapers, lists prepared by community groups, and letters from individuals who claimed they were not enumerated. A number of forms were not usable because addresses or names were missing. Delays were encountered in obtaining address registers because of the priority of other operations using them. The imminence of State closeout dates and the large number of forms to be processed made allocations difficult to complete.

A total of 48,453 housing units were added from the ICR's and "Were You Counted?" forms to the counts for 34 States outside the South. With a figure of slightly over three persons per household used to determine the population to be added, 152,974 persons, or about 0.13 percent of the total population in these States, were added to the census in this operation. Another 12,345 housing units, containing approximately 40,000 persons, were identified for the 16 Southern States. Inasmuch as these missed units might have been duplicated in the post-enumeration Post Office check (see below) in these States, the number of potential adds from ICR's and "Were You Counted?" forms for each ED was compared with the number found in the PEPOC, and the higher of the two counts was used.

## Post-Enumeration Post Office Check

### Introduction

The Bureau's experience in testing its plans for the 1970 census had shown that appreciable gains in coverage could be made by submitting the census list of addresses to a postal review. (See chapter 2 for background.) In mail areas in 1970, a Post Office check of residential addresses was an integral part of the census operation and was carried out prior to Census Day on the mailing list to be used for the enumeration. The fact that no such precensus list existed for those areas of the Nation which were to be enumerated by conventional methods meant that the same type of check could not be made. Therefore, procedures were developed to provide for a post-enumeration Post Office check (PEPOC) in conventional areas in the 1970 census. Since budget restrictions imposed limitations on this program, a PEPOC operation limited to the conventional enumeration areas of the 16 States classified as the South region was incorporated into census procedures. (An analysis of the households missed in the 1960 census had indicated that housing-unit coverage problems were concentrated most heavily in the South.) This major operation added almost 500,000 persons in approximately 170,000 housing units to the conventional area population in these States, an in-

crease of approximately 1.4 percent in the number of inhabitants.

Operationally, PEPOC required the coordinated efforts of the Bureau's district and regional offices, the post offices serving the designated areas, and Bureau staffs in Jeffersonville and Suitland. Overall, the operation consisted of (1) a postal review of the addresses of all housing units enumerated in the conventional areas to identify and report any units potentially missed in the census, (2) a check of these reported units against census records to confirm that they were actually missed, and to eliminate cases definitely enumerated in the census, (3) personal interviews at a sample of the missed housing units to obtain census data representative of this group, and (4) the addition of the appropriate number of units to the census, each being randomly assigned population and housing characteristics obtained from the personal interview sample. The field and Post Office phases of the operation, described in more detail in chapters 4 and 5, are outlined below.

1. District office phase.--As each enumerator visited a housing unit, he prepared a form D-160, "Address Card for Postal Check" (called the "white card"), on which he recorded the address and name of household head or a vacancy description. These cards were then sorted by post office and forwarded to the appropriate post offices for the check operation. Approximately 15 million white cards were generated.

2. Post Office phase.--The local post offices that served the conventional areas in the South were provided with instructions to inspect the white cards and to prepare a form D-711, "Report of Residence Missed by Census" (called a "blue card"), for each housing unit believed to have been missed. (See fig. G.) To assist in determining geography of the suspected missed units, the postmen were instructed to indicate on each blue card the enumeration district and serial

number from the white card for a neighboring unit. (Although the missed unit might have been located physically outside the boundaries of the ED for the neighboring unit, this designation provided a starting point for further investigation.) The post offices then sent their blue cards and any missent white cards, or white cards for undeliverable addresses, to the Bureau's Jeffersonville facility. Approximately 700,000 blue cards were generated. (But, see p.29.)

**Jeffersonville Procedures**

PEPOC processing was conducted between June and September 1970 by a clerical staff which averaged 65 persons.

Upon their receipt in Jeffersonville, the blue cards were sorted by State and by county within State. Within the county classification, the cards again were sorted by ED. ED's then were designated as "high-miss" if (1) at least 15 blue cards were submitted for that ED and (2) these blue cards represented the equivalent of at least 12 percent of the census listings for that ED. All other ED's were classified as "low-miss."

High-miss vs. low-miss ED's.--A great deal of the resources provided for the PEPOC operation were focused on those ED's in which many misses were reported by the Post Office. It was assumed that in such cases either the enumeration was very poor or, for some reason, the postman did not receive the white cards for his route or was unable to make proper use of them, e.g., he might have been unable to identify the individual housing units and therefore prepared blue cards in their place. Thus a case-by-case search of census records was planned for the entire "high-miss" group.

The majority of the missed cases were expected to be in the much more numerous low-miss ED's--the ED's in which the census was relatively complete and the

**Figure G. Form D-711, Report of Residence Missed by Census**

1. ED No. <i>215</i> <i>21</i>		2. Serial No. <i>338</i>	I did not receive a Form D-160 for the family whose name and address I have written below. ◀ These entries were copied from a D-160 for a nearby address.	FORM D-711 (1-21-70) USCOMM-DC <b>REPORT OF RESIDENCE MISSED BY CENSUS</b> 19th Decennial Census - 1970 U. S. DEPARTMENT OF COMMERCE BUREAU OF THE CENSUS
<b>CENSUS USE ONLY</b>				
a.	b.	c.	Family name:	
<i>John</i>		<i>J.</i>	<i>Doe</i>	
First		Initial	Last	
d.	e.	f.	Address	
Route No. and Carrier Initials				
<i>C-9</i> <i>RM</i>			Post Office and ZIP code	
			<i>Sumville, Ga. 50262</i>	

Post Office reported few, if any, misses. For such ED's a sample of cases was selected for intensive review against census records. Data derived from the sample were to be applied to all low-miss ED's.

It was also felt, however, that the characteristics of these low-miss ED households might differ from those of high-miss ED households. Whereas the high-miss ED households were likely to be comparable to the other households in their ED, the low-miss ED households in specific areas--households the enumerator, doing a relatively careful job, had not uncovered--might be more like each other than like already-enumerated households. Separate processing of low-miss and high-miss ED's thus was to be maintained throughout, and separate imputation data (to be used for the missed households to be added to the census) would be obtained for each group. Separate samples were selected from the high-miss and low-miss ED's, each sample to serve as the source of separate sets of imputation data.

Sample selection and sample search of census records.--A sample of approximately 6,000 cases was selected from the blue cards in low-miss ED's. A separate sample of approximately 11,000 cases was selected from the cards in high-miss ED's. (Equal-sized samples of 6,000 cases each had been planned, but the volume of high-miss ED's was such that the sampling patterns provided resulted in almost double the expected number of cases.) Overall, the sample rates were 1 in 60 blue cards for the low-miss ED's and 1 in 13-1/3 (i.e., 3 in 40) blue cards for the high-miss ED's.

Both sample groups were handled as follows: The housing units were searched for in the census address registers in order to determine the extent to which postal reports duplicated units that already had been enumerated. (Attempts to match blue cards to individual questionnaires in the ED cartons proved uneconomical in terms of time and effort, and this practice was discontinued in August 1970.) The number of cases found to be duplicated in each sample was recorded; the remaining cases--approximately 4,000 in the low-miss sample and 6,500 in the high-miss sample--were designated for personal visits.

Both samples were selected from counties which were represented in the Bureau's Current Population Survey (CPS). Such counties not only provided a representative sample of the South as a region but also were regularly visited by a trained staff of enumerators in the monthly CPS. It was assumed that the field interview work required in the PEPOC would benefit from the use of such a staff.

Field samples.--There were two purposes for the field interviews: (1) to provide population and housing characteristics to be assigned to the missed units being added to the census and (2) to determine whether the duplication search of the census records had been adequate. Form D-51 census questionnaires were addressed for each of these cases and mailed to the regional office responsible for the CPS in the particular area covering that address. (The D-51 questionnaire was identical to the D-1 short (100-percent) questionnaire used in the mail-out/mail-back enumeration, but the D-51 included a notice to the respondent explaining that his household presumably was missed in the census.)

CPS interviewers visited each address and either obtained the 100-percent census information for the unit if the respondent believed that it had not been enumerated before or, if it had, so annotated the cover of the questionnaire. The completed questionnaires were returned to Jeffersonville. Approximately 50 percent of the housing units were reported as already enumerated in the census.

Questionnaire mail-out.--For all unmatched blue cards in high-miss ED's remaining after the selection of the field interview sample, labels were prepared and affixed to form D-51 questionnaires. These were mailed directly to the households in question together with return envelopes. Each respondent was asked to (1) complete the questionnaire or (2) indicate on it that the household had been enumerated previously at this or at another address, and to return the questionnaire within 3 days. (Processing deadlines prevented the use of these returns in the PEPOC operation.)

All unmatched blue cards in low-miss ED's remaining after the selection of the field interview sample were sent to one of the Bureau's regional offices, where a form D-26 "Were You Counted?" questionnaire was mailed to each housing unit in question, to be returned only if the units and its occupants had not been enumerated. It was planned that any D-26's returned for the PEPOC operation would be boxed and stored for possible future use. In practice, the form D-26 was identical to the form used in the "Were You Counted?" operation (see p.00), and an unknown number of completed forms D-26 received in Jeffersonville and intended for the PEPOC were processed in the "Were You Counted?" operation instead.

Processing the nonsample cases.--The census records for low-miss ED's were reviewed on a sample basis between June and October 1970. The proportion of units in the low-miss sample found to have been enumerated during the sample search was calculated. Random deletions of all low-miss ED units reported on blue cards by the Post Office then were made at this rate (approximately 30 percent). The actual number of deletions resulting in each ED by this procedure was recorded on a control list. (The control list was prepared on the high-speed printer from a computer tape stripped from the Bureau's master reference file. All conventional ED's were listed for the 16 PEPOC States in State, county, and ED order, and columns were provided for recording all the information necessary to calculate (1) whether the ED would be classified as "high-miss" or "low-miss" and (2) how many adds, if any, would have to be made to the census.)

All nonsample cases in high-miss ED's were compared with the address registers. The search indicated that many of the high-miss ED's actually contained few or no addresses that were missed in the census. All cases found in the address registers (about 44 percent of the total) were marked as duplicates, and the counts were recorded by ED. D-51 questionnaires that respondents returned by mail marked "already enumerated" were discarded together with the corresponding blue cards. The remaining cases were designated as "census adds" and were recorded on the control list.

Preparation of data for additions to the census.--While awaiting the returns from the interviewers, the PEPOC

clerks transcribed the number of missed housing units from the control list onto a FOSDIC-readable breaker sheet for each ED, both low- and high-miss. (These breaker sheets were prepared on the high-speed printer from the same tape that was used to prepare the ED control list.) The population and housing characteristics for the households to be added were obtained from the sample of interviewed households--one set of data for low-miss ED's and another set for high-miss ED's. Each unit added was randomly assigned the characteristics of one of the interviewed households by means of a "cold deck" imputation technique. This was done in two ways for the PEPOC operation:

1. Low-miss cold deck.--The data collected from the low-miss ED interview sample were reviewed in Jeffersonville, and the questionnaires were coded as necessary for input to the FOSDIC operation and microfilmed by State. Questionnaires for households that were interviewed and reported they had already been enumerated were so coded, thus that fact could be "read" in the FOSDIC operation; this coding eliminated the need to alter the number of adds recorded on the breaker sheets in order to conform with the information gained from the field sample. Thus, when a particular household to be added was randomly assigned the data from a questionnaire in the cold deck coded as "already enumerated," the unit was not added. The non-add rate based on the field returns (about 50 percent) decreased the total number of all low-miss ED cases added to about 35 percent. The low-miss decks, one for each of the 16 PEPOC States, were combined into one PEPOC low-miss deck.

2. High-miss cold deck.--The field data for the high-miss ED sample also were reviewed. The questionnaires for households claiming to have been enumerated were discarded. As all "previously enumerated" cases had been eliminated during the earlier processing of the high-miss ED cases, no further adjustment of the deck was considered necessary. (The non-adds for high-miss ED cases were not processed on a sample basis as the low-miss cases had been. The cold-deck adjustment for low-miss ED's was required because the sample results were to be generalized in all cases.) The high-miss questionnaires were processed to create decks, and the decks for the 16 States were combined into one PEPOC high-miss deck.

For each ED on the PEPOC control list, the specified number of housing units was selected in sequence from the appropriate cold deck. The housing information and the population characteristics for each member of the household were duplicated and added to the ED. If the ED was tabulated by block, the PEPOC additions were distributed among the blocks.

Modification of the PEPOC procedures.--Approximately 500,000 blue cards were expected from the post offices, but more than 700,000 were received. This excess was due in part to irregular timing in the transmittal of white cards to the post offices, shipment of some to the wrong post offices (estimated at 100,000 of the 15 million white cards), and errors in the postal addresses on some cards. A procedure was added, therefore, to match missent white cards from one post office to the blue cards from the post office that should have received

the missent cards; this eliminated about 5 percent of the volume prior to any processing. Where post offices had not sent in any blue cards, but the white cards destined for them had been misdirected, the missent white cards were redirected to them so the post office check could be completed.

It was found during the course of processing that about 25 percent of the blue cards lacked ED identification. These cases required allocation to ED's. Not only did this slow down processing and increase the cost of the operation, but the ambiguities in allocation reduced the usefulness of the classification of "high-miss" and "low-miss" ED's.

To enable the PEPOC operation to keep up with the Bureau's time schedule for processing the 100-percent census data, changes were made in the PEPOC procedures in August 1970 before first-count processing began for any of the States. These changes were as follows:

1. A simplified ED allocation procedure was instituted so that less time was spent examining detailed maps.

2. The processing of questionnaires mailed back from respondents was discontinued.

3. The high-miss and low-miss field samples were combined. The address registers were searched for all cases allocated to high-miss ED's, but the characteristics to be added were taken from the one data deck which now included units to be deleted as well as those to be added. This increased the deletion rate in high-miss ED's to a total of about 70 percent.

4. Additional matches were instituted for all postal city-delivery ED's in which a large number of unmatched blue cards remained after the initial search of the census records. This involved review of the geographic allocation of the cards as well as of the matching operation itself. Also, all field sample returns reported as "not previously enumerated" were matched to the address registers (and reallocated if necessary) to make certain that no duplication had occurred. The rematch in city-delivery ED's did increase the match rate, but the review of the field cases produced practically no changes.

5. To simplify the overall census processing, the breaker sheet entries of potential adds from the PEPOC operation were adjusted to incorporate the potential adds from the supplemental forms operation (i.e., the "Were You Counted?" forms, late-arriving individual census reports, etc.; see p. 26). ED-by-ED comparisons between the potential adds from PEPOC and potential adds from the supplemental forms were made, and the larger of the two counts was indicated on the breaker sheet.

The PEPOC procedures were modified once more after all the first-count census processing was completed. When the data were reviewed, it was found that the race distribution was erroneous for 13 of the 16 Southern States. The assignment of household characteristics to the added housing units in PEPOC had resulted in the imputation of persons of one race in counties whose inhabitants were known to be virtually all members of

another race. To correct for this error, an alternate imputation procedure was employed for 13 of the 16 States, (Delaware, Mississippi, and South Carolina were found not to require correction.) Under this procedure the total population counts, including PEPOC adds, were held constant, and the characteristics for the households to be added were imputed from other households in the ED in which the adds were made.

### National Vacancy Check

In the analysis of the undercount following the 1960 census and subsequent tests of alternative census procedures, the Bureau discovered that a significant source of undercount was the census enumerators' classification of some housing units as "vacant" when in reality they were occupied.

Between June and October 1970, as soon as completed address registers were available, clerks in Jeffersonville selected a national sample of 15,000 housing units that had been enumerated as "nonseasonal vacant" to be revisited. The random, self-weighting sample was selected in a subset of the counties currently included in the Bureau's CPS. Within each county (or group of counties) a systematic sample of ED's was selected, and half of the nonseasonal vacant units in these ED's were designated for interview. About 1,500 ED's were included in the sample, with an average of about 10 units per ED.

CPS interviewers visited each of these sample units and interviewed the current residents (or neighbors, if the unit was vacant). The purpose of the interview was to determine (a) the occupancy status of the unit during the entire period from April 1 to the time of the interview, (b) whether the sample housing unit was the usual place of residence of the occupants, and (c) the number of occupants of the unit. Replies were entered on form D-910, "Record of Occupants."

In Jeffersonville the completed forms were coded either as occupied at the time of the census or as vacant. Since the rules for census enumeration provided that the occupancy status was to reflect the situation on the day of the enumerator's visit, and since the enumeration days were not known exactly, some ambiguity existed for those units which were vacant during part of the census period. (The census period in conventional areas was approximately April 1 to May 31, and in mail areas April 1 to July 31.) A conservative rule for coding these cases was established: Any unit that was not continuously occupied by a single household during the entire census period, and was enumerated as vacant in the census, was considered to have been classified correctly. A further step was taken to ensure that the error rates were not overstated. For all sample units coded as occupied, checks were made to guard against the possibility that the occupants of the wrong unit had been interviewed because of poor address identification. Whenever the sample unit was in a multiunit structure, the occupant's name was compared with the names reported in the census in all other units in the building. For one-family homes, if the address was not a specific city-type address, the name was compared with the names of occupants of a group of homes shown in the census as located on either side of the sample unit. A small number of errors were detected and corrected by this device.

Clerks tabulated the results both nationally and for 12 geographic divisions. Separate tabulations were made for the Northeast, North Central, South, and West regions of the United States and, within each region, for areas which were primarily urban, primarily rural, and mixed (urban and rural). For each area two estimates were made: the proportion of misclassified units and the average number of persons per misclassified unit. A third, derived figure also was calculated for each of the 12 geographic divisions, consisting of the proportion of misclassified units multiplied by the ratio of the average number of persons per misclassified unit (2.36) to the U.S. average number of persons per household (3.1). Thus, in all 12 divisions the ratios were less than 1.0 because the misclassified housing units tended to be occupied by smaller-than-average households.

The results then were used in the computer processing of the 100-percent edit. In the computer a systematic selection was made of every  $n^{\text{th}}$  nonseasonal vacant housing unit in each ED, and it was mechanically converted to an occupied unit. The number of occupants in the preceding occupied housing unit, together with their population characteristics, was inserted into the record for the converted unit. (Housing characteristics were supplied in the editing and allocation procedure used for all housing units.) The value of  $\bar{n}$  (i.e., the rate of conversions) conformed to the third set of estimates described above, that is, the vacant-housing-unit error rate adjusted to reflect the smaller households in misclassified units.

Two minor modifications in this procedure sometimes were applied. The first was that no more than 25 conversions per ED were permitted, to guard against unusual situations. Only a few ED's were affected by this rule. The second modification was caused by the necessity to process some States for the census before all the field work and final tabulations of the vacancy check were completed in the latter part of October 1970. In these cases preliminary estimates of the required rates, based on the regional returns tabulated at the time a State was to be processed, were used.

The status of 380,209 housing units was converted from vacant to occupied, representing 8.3 percent of all nonseasonal vacant units, and 1,068,882 persons (about 0.5 percent of the enumerated population) were added as a result of this operation. (In order not to overstate the resultant number of occupied units while accounting for the correct number of persons, the value of  $\bar{n}$  was adjusted so that the average household size for the housing units converted was 2.83 persons.)

### Count Adjustment

There were two stages to this operation, which was initiated in June 1970 and carried on as necessary thereafter. The first stage was in response to complaints arising from the announcement of the preliminary counts, the second to the final published figures at block-group (ED) or block and tract level. There were as many as 26 staff members employed at Jeffersonville at the height of these activities. The purpose of this operation was to search the appropriate address registers to determine whether or not households were missed or miscoded geographically. (For results, see sections on "Under-

count Complaints" and "First-Count Data Review" in chapter 5.)

In the first stage, register search was required for over 700 of the inquiries or complaints which the Census Bureau received from local officials after preliminary population and housing counts had been announced and the district offices had been closed. When a specific address was reported, the address registers for the ED involved as well as for the surrounding area were checked to see if the reported household was already included. If the address was not found anywhere, the household and housing unit were considered as missed and were added to the census by means of a supplemental questionnaire. If addresses were found in areas outside their proper ED's, the questionnaires were moved from one ED to another, and the affected ED's were remicrofilmed and reprocessed on the computer. If 100-percent census processing had been completed for a State, the questionnaires could no longer be removed, so a correction note was published in the report(s) for that State.

Complaints arising from the final published figures were handled in the second stage of the operation: Bureau personnel in Suitland reviewed all known source materials for each ED in question to determine whether or not there had been a processing error. If not, clerks in Jeffersonville plotted on maps the addresses shown in the address registers; those addresses which fell outside the ED boundaries were transferred to other ED's, and adjustments were made for any spots which appeared to have been missed. Correction notes were prepared for the appropriate publications; the tabulations in the printed reports themselves were not revised.

## COMPUTER PROCESSING

### Equipment

The Bureau used electronic computers for the first time in processing the 1950 decennial censuses of population and housing. The UNIVAC I (Universal Automatic Computer) equipment purchased in 1951 for that purpose was retired in 1963. The larger UNIVAC 1105's purchased between 1958 and 1962, and used for the 1960 decennial and other censuses, were replaced by later 1100 series UNIVAC systems on which the 1970 census data were processed. The UNIVAC 1107, 1108, and 1106 systems, together with the IBM 1401 computer system and some of the other peripheral equipment used in 1970 census processing, are described below. (The Bureau's FOSDIC system is discussed on pp. 7-9.)

Two UNIVAC 1107 thin-film computers were purchased in 1963 and 1964 to replace four of the five UNIVAC 1105 computers. (The Bureau's last 1105 computer was retired in 1965.) The UNIVAC 1107 was a high-speed, large-scale electronic binary computer with a repertoire of 117 different instructions for performing various arithmetical, logical, and control functions. This computer could accomplish the operations of several entirely different programs simultaneously. Under this system the central computer could continue internal work during input-output operations. Each model 1107 computer had the following components:

1. 18 Uniservo tape handlers--4 Uniservo II's with reading and writing rates of 20,000 binary-coded alphanumeric characters per second (these handlers were compatible with the 1105 equipment), and 14 Uniservo IIIA's with reading and writing rates of 100,000 characters per second.

2. A memory composed of 65,536 words of core storage--approximately eight times the capacity of the 1105--as well as 128 words of thin-film instruction storage. (Thin-film storage is capable of data transfers at speeds measured in billionths of a second.)

3. Two magnetic drums, each with a capacity of 786,532 words (each word containing 6 characters, or 36 bits), or nearly 100 times the capacity of the model 1105 computer. The 1107 drums were each equipped with 880 read/write heads.

4. One on-line punchcard reader, capable of reading up to 600 cards per minute.

5. One on-line high-speed printer, capable of printing 132 characters per line at the rate of 700 to 922 lines per minute (depending on the number of characters per line being printed).

6. One on-line paper tape reader punch subsystem, used for minor amounts of input and output.

In November 1967 the Bureau acquired its first UNIVAC 1108 computer, and a second was installed in March 1969. The two UNIVAC 1107 computers were converted to 1106's in the winter of 1971-72. The 1106 and 1108 systems were basically the same as the 1107 computer described above, except that they had more sophisticated internal circuitry and had under gone certain modifications in the ways that data were stored. The essential differences among the systems were in their speed of operation: The 1107 had a basic memory cycle time of 4 microseconds (millionths of a second); for the 1106 this was reduced to 1.5 microseconds; the 1108 computers had a cycle time of 750 nanoseconds (billionths of a second).

Until mid-1972 these computers were controlled by a master control program known as EXEC 1. This program was loaded into the UNIVAC computer system prior to starting any processing. It communicated by using on-line teleprinters to type out various messages that kept the operations personnel informed of the progress of the running programs. The computer operator's console also included an on-line typewriter for communicating with the executive system. (EXEC 1, which supported only a small number of computer programming languages, was replaced by a new master control program, known as EXEC 8, which became operational in mid-1972. All of the 1970 census tabulation work, however, was completed under the EXEC 1 system.)

Some of the 1970 census processing operations, although complex enough to be run on a computer, did not require the use of a large-scale system such as the 1100 series UNIVAC's. The Bureau acquired two IBM 1401 computer systems in 1962: One was a magnetic tape system used for the translation of machine language

between two otherwise incompatible systems, IBM and UNIVAC, and for the preparation of data for input into the larger UNIVAC system; the other IBM 1401 was a card system used as an extension of the electrical and mechanical card processing systems used in the Bureau and was employed for small programs such as payroll, progress reporting, inventory control, etc. The card system was capable of reading 800 cards per minute and printing out 600 lines of data through an on-line printer.

The Bureau had four high-speed off-line printers capable of printing 400, 600, and 720 lines per minute. Input to these printers consisted of magnetic tape produced by any of the computers. The magnetic tape was read by a tape handler similar to those serving the computers. The handler transferred the data to a small magnetic core buffer memory. In the printer itself there was a revolving drum made up of 130 type wheels, each containing 51 different letters, numbers, and special characters. UNIVAC codes for enough characters to print one line were transferred from the buffer memory to a vacuum-tube printer memory. The contents of the printer memory were so interpreted that each type wheel reached the position in which the proper character was in place behind the printing ribbon and the paper being fed through the machine. Hammers then were activated which pressed the paper and ribbon against the characters on the wheels. The paper then was advanced one line and the process was repeated as necessary, all at the rate of 10 lines per second. The printed output of the computer results thus could be reviewed clerically or, when produced at the slower 400-lines-per-minute speed, used directly for offset printing of census publications. (See chapter 10.)

For reference and storage purposes, however, it was more practical--and considerably less expensive in terms of equipment and storage space--to have the printed-out data on microfilm rather than on bulky stacks of paper. In 1967 the Bureau began using an SC 4411 computer document recorder. This device accepted digital signals from magnetic computer tape and converted them into alphanumeric images directly on 16 mm microfilm without the need of intervening paper and printing. This was done at speeds 30 to 40 times faster than the high-speed printer. Prints for individual use or for offset printing could be made from the microfilm records if needed. The recorder also had the ability to convert certain computer codes into retrieval code patterns that could be inserted between frames of data, thus facilitating document location.

The MIRACODE (microfilm information retrieval code) system, developed in 1967, made it possible to locate any document from a file of nearly a million pages on microfilm, and to make a photo facsimile of it, within 30 seconds. The heart of the MIRACODE system was a retrieval station that stored up to 490 film magazines, a reader that searched at high speed for the retrieval codes on the film, and a dry-process film printer for producing enlarged paper copies of the microfilm frames.

### Staffing

During the decennial census period there were 20 to 30 programmers and five to eight support persons engaged in

preparing the 43 major programs for processing the 1970 census data. Approximately 15,000 man-days, including about 3,700 for supervision, were required for this purpose.

There were approximately 600 persons engaged in all other phases of the Bureau's computer operations. This staff handled the computer processing of all of the Bureau's censuses and surveys as necessary, so that few of these persons were involved exclusively with the 1970 census. There were times, however, when virtually all efforts were concentrated on 1970 census processing. The personal time of this staff was charged to the computer operations; the various censuses and surveys then were charged in terms of computer hours, so that the number of man-days expended was not calculated. For the number of computer hours and costs charged to decennial census processing projects, see p. 40.

### Computer Processing of 100-Percent Data

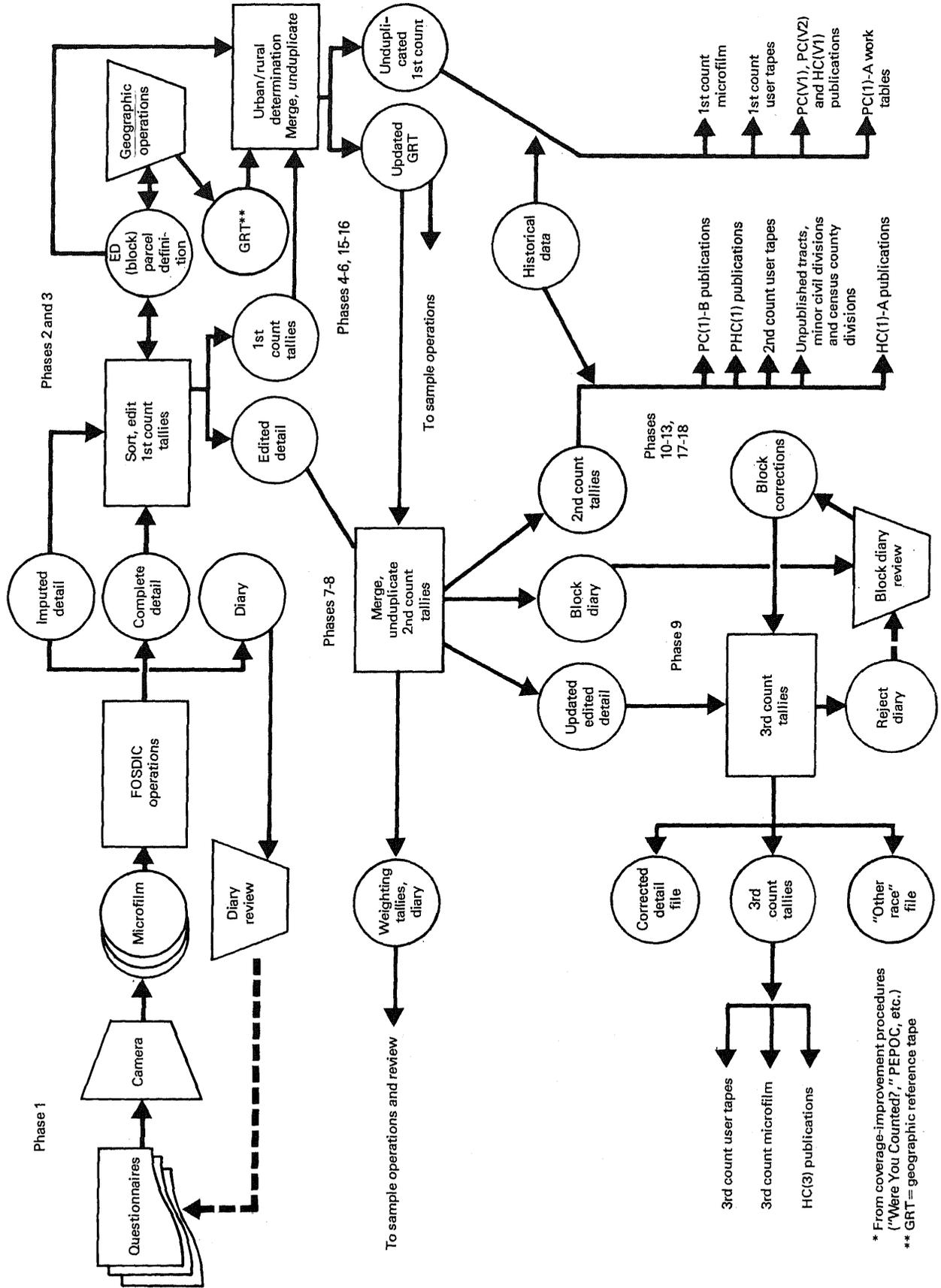
Introduction.--The basic processing of the 100-percent data (those collected for all persons and housing units), beginning with microfilming through the various computer phases, is illustrated in an abbreviated flow chart, figure H, and the timing is shown in table 1. Much of the system used was developed during the dress rehearsals for the 1970 census, but additions and modifications were made to that system for use in the decennial census processing.

The FOSDIC operation, described on pp. 7-9, was known as Phase 1 of the 100-percent data-processing operation. The computer processing for the various coverage improvement procedures (the "Were You Counted?" operation, the post-enumeration Post Office check, etc.) is outlined under their respective headings; see p. 26ff. Subsequent phases, which were numbered as they were planned, are described below in the order in which they were performed.

Sort, edit, and first-count tally--Phases 2 and 3.--Phase 2 involved the combining of 10 to 15 FOSDIC tapes for a State (or from 300 to 500 ED's) into one sort pass, i.e., all the questionnaire records were put into their proper sequence. The sort sequence was by State, county, ED, descending FOSDIC work unit, address serial number, and type of questionnaire (first form, continuation, vacant, etc.). Phase 2 contained a recovery system to allow data to be bypassed when an unrecoverable "read" error occurred. The affected data were reviewed clerically and recycled as needed.

Phase 3 involved merging from one to three sorted tapes to consolidate the record file, then editing it and tabulating the first-count tally. The edit eliminated inconsistencies and allocated missing entries for both population and housing items and also introduced data from the National Vacancy Survey. The first-count tallies were produced simultaneously in three different formats, depending on the type of ED and area: For conventional enumeration areas the lowest-level tallies were at the ED level; for mail enumeration areas, tallies also were by ED when the address register was compiled by prelisting, but by block group within ED when a computer-generated address register was used; for any mail ED which was identified as located in a potential urban fringe area (i.e., an area which might

Figure H. Simplified Flow Chart of 100-Percent Data Processing



\* From coverage-improvement procedures ("Were You Counted?", PEPC, etc.)  
 \*\* GRT = geographic reference tape

have sufficient population density to warrant the tabulation of its statistics by block), the tallies were by block within ED.

For the potential urban fringe area, these block summaries, along with predesignated ED summaries (defined in the ED parcel definition tape), were used to define the final urban fringe zones of urbanized areas.

Different parts of the standard computer programs--or variations of them--were used to correct FOSDIC errors, to process States when block parcel tapes for them were not available, and to handle tabulations for a State being processed after its population totals were fixed.

Merge and unduplication, geographic updating, summarization of first count, advance and PC(1)-A final reports--Phases 4-6, 15-16.--In Phase 4 the first-count tallies, in summary form, were merged and unduplicated (i.e., duplicate records were identified and removed) and matched to the geographic reference tape. This tape contained the identification numbers of the ED's with most of their associated geographic codes. The match produced a listing of all problems, such as unmatched summaries, unmatched GRT cases, duplicates, etc., which then were handled by recycling missing ED's through the computer or by updating the GRT. This phase also provided population sizes for towns in New England and for places elsewhere, along with advance population and housing counts for each State, its counties, and its places.

In Phase 5 an updated ED (block) parcel definition tape was produced. This tape was used to define the final urban/rural status of the first-count tallies and the ED records on the GRT for the potential urban fringe. Urbanized-area codes and place-size codes were assigned both to the tallies and to the GRT records. Codes for unincorporated places failing to meet the size requirements were eliminated from the data tape and from the GRT. A disposition listing was produced which identified those urban/rural definition problems which would have to be reviewed clerically and resolved. (For details of (block) parcel tape preparation and use, see chapter 3.)

Phase 6 used the updated first-count tallies to produce two different summary tapes--ED summaries for conventional and prelist ED's and block-group summaries for tape address register ED's. The second summation created the levels required for the advance reports (see below), for the series PC(1)-A reports, and for analytical tables. This summarization--approximately 400 cells of complete-count data in 55 tally matrices for each of about 250,000 ED's (File A), and for larger areas up to the State level (File B)--also was made available on tape and microfilm for users of census data.

In Phase 15 (which followed Phase 6 in the operational order) the tables were developed for the population and housing advance reports (series PC(V1), Final Population Counts; PC(V2), General Population Characteristics; and HC(V1), General Housing Characteristics). Work tables also were developed for States, as were selected tables to be used in the PC(1)-A series, Number of Inhabitants, and analytical tables for State and lower-level geographic areas. This was done by matching the master reference

file (MRF) name tape and the historical data produced in Phase 17 to the first count files created in Phases 5 and 6. A table program was used to expand, convert, and format the binary data on the tapes into the required table formats and to compute the necessary percentages, medians, and ratios. The appropriate routines to prevent disclosure were applied during formatting.

The output of State data in Phase 15 was summarized in Phase 16 by regions and divisions of the United States, and for the Nation, to produce the U.S. summary report.

Detail record merge and second count--Phases 7 and 8.--Phase 7 was a custom-coded merge program in which the detail record files from Phase 3 for each State or quasi-State (a portion of a State handled for processing purposes as if it were a complete State) were merged by their breaker records only. When two or more breakers contained the same identification (State, county, and ED number), the breaker with the highest FOSDIC work unit number (representing the latest processing of the particular ED) was retained together with all of its data. The other breaker records, with their associated data, were dropped from the file.

Phase 8 had as input the merged detail file for each State as received from Phase 7, and the updated GRT for the State which was created in Phase 5. The program provided the following four output files: (1) Block diary, (2) second-count tallies, (3) updated, edited detail file, and (4) sample weighting tallies.

To produce these, the following computer actions took place: The merged detail file and the GRT were matched, so that each ED would have its proper geographic identification. ED's in contract block areas (those for which local authorities had contracted with the Census Bureau to provide statistics at the block level) were identified, and the data were classified as either rural or urban. In the ED's that were tabulated by block, the file was edited to make certain that all data were assigned to blocks acceptable for that ED. A diary of this action was produced for clerical review. (For further details, see p. 13.)

After the above, the updated, edited detail file was produced, as well as second-count tallies which contained approximately 3,500 cells of complete-count data in 93 tally matrices for each tract or pseudo-tract on the census basic record tapes. Separate tally matrices were maintained for the urban and rural portions of the summary segments. Sample weighting tallies were produced for use in later sample processing.

A sample analyzer program was used to perform checks on the sample records contained in the detail file. The actual number of housing units in the 20-percent sample was compared with the expected number, which was based on the number of housing units enumerated on complete-count questionnaires. The number of persons in the sample was compared with the expected number, which was based on the 100-percent count of persons and housing units. The results of these checks were displayed for clerical review.

Summarization of the second count (population and housing, United States and States), tract summarization,

and historical data--Phases 10-13, 17-18.--Summarizations of the second-count tally file were produced in six phases to supply (a) input for the Linotron table preparation program leading to publication, (b) data for unpublished tract summaries for about 100 counties and places outside SMSA's, and (c) second-count summary tapes for census users after necessary reformatting. The phases were as follows:

Phase 10, for the series PHC(1)-B publications (General Population Characteristics), covered the geographic level summarizations for States, SMSA's (including those crossing State lines), counties, urbanized areas, towns and places of selected sizes, and by urban and rural classification. Housing data from Phase 12 (see below) were merged with population data to prepare the second-count user tapes, files B and C. Work producing the U.S.-level data was accomplished in Phase 13.

Phase 11 was for the PHC(1) series publications (Census Tracts). Here the summarizations were by tract for 241 SMSA's in the United States and Puerto Rico. Control files were printed out for the assistance of clerical personnel who were forming the page images for the published tract reports. The second-count user tapes, file A, were produced here, as well as tract tables for the nonmetropolitan tracted areas. Optional tract-type tables for minor civil divisions and census county divisions also were produced.

Phases 12 and 18 were concerned with the series HC(1)-A publications (General Housing Characteristics). The procedures and geographic-level summarizations paralleled those in Phases 10 and 13 above.

Phase 17 involved the preparation of the historical population and housing data used in Phases 10, 12, and 15.

Third count--Phase 9--Activities were varied. One program in Phase 9 was used to prepare the block number corrections produced in the block statistics diary operation (see p. 13) for later correction of the detail file. Another was used to produce the third count (block tallies) itself, to insert the block number corrections, to assign contract block identification to designated block data summaries, and to produce an "Other Race" detail file. Later programs were used to summarize the block summaries to the levels required to produce the regular publications of block statistics and the contract block reports, to make necessary calculations (e.g., computing averages), and to extract data from the summaries for input to the Linotron table program that was employed to produce the published reports. The regular block reports at the urbanized-area level contained tallies for places of 2,500 inhabitants or more, and for tract and blocks. Also produced were user tapes and optional microfilm displays of the third-count tallies. These tallies contained approximately 250 cells of complete-count data in 38 tally matrices for each of about 1.7 million census blocks in the United States and Puerto Rico.

Phase 14--The operations originally planned for Phase 14 were subsumed in the other phases.

## Computer Processing of Sample Data

The processing of the sample data, beginning with microfilming through the various computer phases, is illustrated in an abbreviated flow chart, figure 1, and the timing of the computer phases is shown in table 1. Much of the system used to process the sample was developed for the dress rehearsals of the 1970 census. Although the dress rehearsal sample data were never published, they nevertheless were processed through the preparation of summary tapes for census users in order to test the various steps in the system. Where practical, additions and modifications were made to the system, and new programs were written as needs became apparent during actual decennial processing.

The FOSDIC operations described on page 9 were known as Phase 100 of the sample data-processing operation. No pre-editing was done, and no diary was produced in this phase. The output of Phase 100 was fed directly into Phase 101. This and subsequent phases of the computer processing are described below. (The phases, which were numbered arbitrarily when they were planned, are described in the order in which they were performed.)

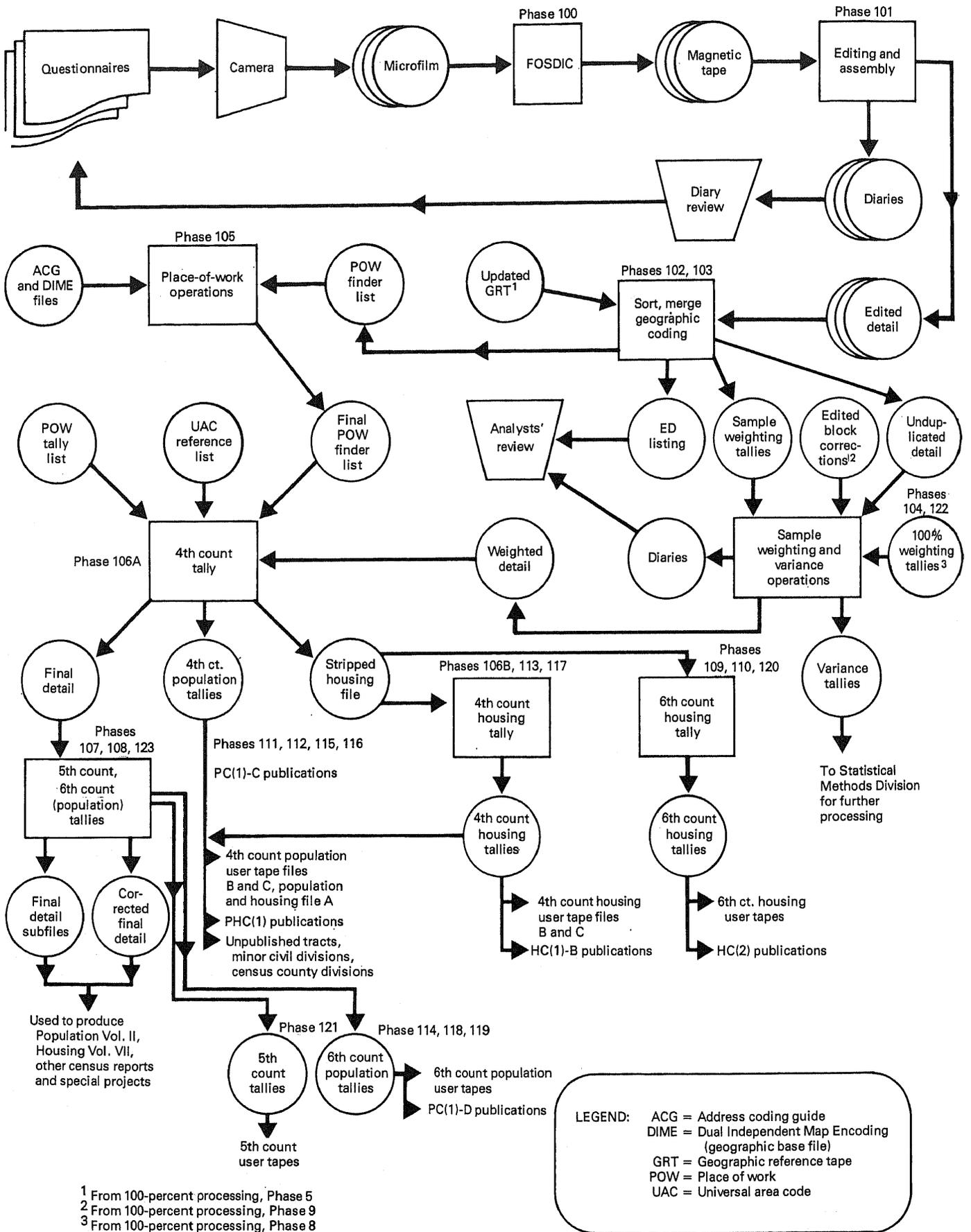
### Development of the Base File

Edit and assembly--Phase 101--About 14 FOSDIC tapes for the same State were combined into a single computer run, producing one full computer tape of edited detail records. (One roll of microfilm contained approximately 3,600 questionnaire pages, and seven full rolls of microfilm--one FOSDIC work unit were converted to one FOSDIC magnetic tape.) The Phase 101 program, like Phase 2 (see p. 32), had a recovery system to allow data to be bypassed when an unrecoverable "read" error occurred. The affected data were reviewed clerically and recycled as needed.

Two outputs resulted from the Phase 101 computer edit: One was a tape that included the detailed data assembled for each household or group quarters, with missing or inconsistent responses allocated. (See chapter 15 for allocation procedures.) The other was a diary of the allocations and counts by sample item for each ED. This diary was used in the diary review operation. (See p. 14.)

Sort, merge, and geographic coding--Phases 102-103--The household and group-quarters records were sorted (put in sequence) by State, county, ED, FOSDIC work unit in descending sequence, and address serial number--the same sort sequence used in the 100-percent files. Next, sorted sample data files were merged into a single file per State (or quasi-State). Duplicate records were removed, and complete geographic identification was assigned to each ED record. A listing was produced, displaying for each ED both the 100-percent and sample population and housing counts. ED's for which there were 100-percent data but no sample data (or vice versa) were identified. A file of workers was produced for use in Phase 105; this file contained the records for persons who were at work during the week prior to enumeration and for whom a 16-digit place-of-work address code had been assigned clerically. (See p. 16.) A file of tallies was

Figure I. Simplified Flow Chart of Sample Data Processing



<sup>1</sup> From 100-percent processing, Phase 5  
<sup>2</sup> From 100-percent processing, Phase 9  
<sup>3</sup> From 100-percent processing, Phase 8

**LEGEND:** ACG = Address coding guide  
 DIME = Dual Independent Map Encoding (geographic base file)  
 GRT = Geographic reference tape  
 POW = Place of work  
 UAC = Universal area code

generated which became the base for producing sample weights in subsequent programs.

**Table 1. Starting and Completion Dates for Principal Computer Processing Operations in the 1970 Census**

(U.S. questionnaires only)

Operation	Phase	Started	Completed
<b>100-PERCENT PROCESSING</b>			
FOSDIC.....	1	May 1970	Dec 1970
Preliminary detailed record area sort.....	2	May 1970	Dec 1970
First count--portfolio edit and tally.....	3	May 1970	Dec 1970
MEDlist (Master Enumeration District List).....	--	Jun 1970	Jun 1971
Geographic identification A.	4	Jul 1970	Jan 1971
Geographic identification B.	5	Jul 1970	Jan 1971
First count--summarization..	6	Jul 1970	Feb 1971
Final detailed record area sort and merge.....	7	Sep 1970	Feb 1971
State PA summarization.....	15	Jul 1970	Mar 1971
U.S. PA summarization.....	16	Dec 1970	Jun 1971
Second count.....	8	Sep 1970	Feb 1971
State PB summarization.....	10	Sep 1970	Oct 1971
U.S. PB summarization.....	13	Jul 1971	Dec 1971
Tract summarization.....	11	Sep 1970	Jan 1972
State HA summarization.....	12	Sep 1970	Oct 1971
U.S. HA summarization.....	18	Jul 1971	Nov 1971
Third count--blocks.....	9	Nov 1970	Jun 1971
Ward summarization.....		suspended	
<b>SAMPLE PROCESSING</b>			
FOSDIC.....	100	Jan 1971	Sep 1971
Edit and assembly.....	101	Jan 1971	Sep 1971
Area sort.....	102	Feb 1971	Sep 1971
Geographic identification and merge.....	103	Mar 1971	Sep 1971
Weighting.....	104	Apr 1971	Dec 1971
Place-of-work match.....	105	Apr 1971	Dec 1971
Fourth count tabulations--population.....	106A	Apr 1971	Jun 1972
Fourth count tabulations--housing.....	106B	Apr 1971	Jun 1972
State HB summarization.....	113	May 1971	Apr 1972
U.S. HB summarization.....	117	Mar 1972	<sup>1</sup> Nov 1972
State PC summarization.....	111	May 1971	Apr 1972
U.S. PC summarization.....	116	Mar 1972	<sup>1</sup> Oct 1972
Tract summarization.....	112	May 1971	Apr 1972
Fifth count.....	107	May 1971	May 1972
ZIP summarization.....	121	Nov 1971	Jun 1972
Sixth count--population....	108	May 1971	Nov 1972
State PD summarization.....	114	May 1971	<sup>1</sup> Nov 1972
U.S. PD summarization.....	119	Jun 1972	<sup>1</sup> Jan 1973
Sixth count--housing.....	109	Jun 1971	Jun 1972
Housing Vol. II State summarization.....	110	Jun 1971	Jun 1972
Housing Vol. II U.S. summarization.....	120	Apr 1972	Jul 1972

-- Never assigned to a specific phase.

<sup>1</sup>Includes farm "redo"; data from the farm "redo" for earlier phases were produced as supplementary reports and are not reflected in this chart.

Sample weighting and variance operations--Phases 104 and 122.--Phase 104 utilized counts produced in Phase 8 of the 100-percent processing (see p. 34) and counts produced in Phase 103 of the sample processing. The two sets of counts were matched in the first Phase

104 program, and a ratio estimation procedure was used to produce weighted values. These values were assigned by the second program in the Phase 104 series to every housing unit or person in the sample detail file, based upon the characteristics of that housing unit or person. Future tabulations of the sample detail file were accumulations of these sample weights. The detail file also was updated in this second program by a match and correction based on the block statistics corrections made during Phase 9 of the 100-percent processing. (See p. 35.) In Phase 122, tallies were prepared from the sample detail file for purposes of measuring the variance in weights produced and assigned in Phase 104.

Place of work--Phase 105.--One of the products of the construction of address coding guides and DIME (Dual Independent Map Encoding) files (described in chapter 3) was a place-of-work reference file. Producing this file from the coding guide computer records, and matching it to the place-of-work finder list produced in Phase 103, was known as Phase 105. The reference file was used by the computer to code place-of-work data according to 16-digit geographic codes.

#### Fourth Count Processing

In the fourth count the data were tabulated and tables were prepared for the Population Series PC(1)-C reports (Volume I, Characteristics of the Population, Chapter C, General Social and Economic Characteristics); Housing Series HC(1)-B reports (Volume I, Housing Characteristics for States, Cities, and Counties, Chapter B, Detailed Housing Characteristics); and Joint Population-Housing Series PHC(1) reports, Census Tracts. User summary tapes also were produced in this process; these tapes contained 127 population matrices with approximately 14,500 cells of data for each area and 200 housing matrices with approximately 12,200 cells of data for each area.

Population PC(1)-C--Phases 106A, 111, 115, 116.--In Phase 106A the detail file from Phase 104 and the place-of-work finder file from Phase 105 were matched to give the detail file its final updating in the place-of-work area. A universal area code file was used to create migration recodes, certain other recodes were established, and a housing detail file was stripped (i.e., selected items were extracted). Tallies were produced at the tract or pseudo-tract level by race and by residence. Phase 111 consisted of certain programs which enabled the computer to--

Select the appropriate sample population records necessary to produce tabulations for tract processing (Phase 112);

Sum the data to higher levels of geography;

Identify and combine summaries for tabulation areas that crossed State boundaries;

Develop fourth-count population summary tapes and user tapes (file B and C) with appropriate suppression to protect confidentiality;

Extract the appropriate data for the publication tables; and

Prepare these tables for the Linotron phototypesetting system.

Phase 115 was a computer operation in which another series of programs was used to develop historical data cells for the publications. In Phase 116 the sample data were summarized at the U.S., region, and division levels for the U.S. summary report.

Housing HC(1)-B--Phases 106B, 113, 117.--The stripped housing file created in Phase 106A was used in Phase 106B to produce housing tallies for the same levels as the population tallies. In Phases 113 and 117 functions similar to those in Phases 111 and 116 were performed.

In an additional computer run in the spring of 1972, special counts were produced for use in correcting a misclassification found in the statistics on rural farm or rural nonfarm population in the Population Series PC(1)-C State reports and the statistics on occupied rural farm and rural nonfarm housing units in the Housing Series HC(1)-B State reports, as well as in comparable statistics reported in both the population and housing files of the fourth-count summary tapes. (The misclassification was limited to the farm/nonfarm components and was compensating between the components. Therefore, the population and housing unit figures for the rural areas as a whole were not affected, nor were the figures for urban areas or for each State as a whole.)

Joint Population-Housing PHC(1)--Phase 112.--Data from Phases 111 and 113 were summarized and prepared as input to the Linotron system for the Census Tracts reports and in the form of summary tapes (file A).

#### **Fifth Count Processing**

PD strip-ZIP block group tallies--Phases 107, 121.--The fifth-count tallies were produced by using the ZIP percentage file, which specified the percentage of housing units covered by a particular postal ZIP code in each ED. These tallies presented approximately 900 cells of sample population and housing data in 53 matrices. The data needed for the sixth-count population tallies were stripped from the final sample detail file and separated into three files by housing unit--an urban file, a rural nonfarm file, and a rural farm file. Vacant units were dropped. Also produced in this phase was an ED/block group file similar to the ED/block group file generated in Phase 3 of the 100-percent processing. (See p. 32.) In Phase 121 the ZIP code tallies produced in Phase 107 were summed to the 3-digit level everywhere, and to the 5-digit level in SMSA's, and were used to prepare ZIP code summary tapes for census users, with appropriate suppression. (The fifth-count tallies appeared only on summary tapes, not in printed reports.)

#### **Sixth Count Processing**

In the sixth count the data from the appropriate stripped file of the basic census records were tabulated and tables were prepared for the reports in Population Series PC(1)-D (Volume I, Characteristics of the Population, Chapter D, Detailed Characteristics) and Housing Series HC(2) (Volume II, Metropolitan Housing Characteristics).

Population PC(1)--Phases 108, 114, 118, 119.--In Phase 108 the tallies were produced from the files

obtained from the fifth-count processing. The tallies were divided among eight tally programs in which over 150,000 data cells were produced for 92 separate matrices. In Phase 114 the data were summarized to the various levels required for the tables; the user summary tapes were prepared; the sixth-count publication tables for the various PC(1)-D reports were extracted, and these tables were prepared in the appropriate format for the Linotron system. In Phase 118 historical data from the 1960 Population Volume I, Chapter D, reports were prepared for use with the comparable 1970 census data. In Phase 119 the sample population data were summarized at the U.S., region, and division levels for the U.S. summary report.

Housing HC(2)--Phase 109, 110, 120.--In Phase 109, sixth-count housing tallies were produced from the stripped housing detail file generated in fourth-count Phase 106A. Phase 109 consisted of six programs in which approximately 110,000 cells of data were produced for 348 separate matrices. In Phases 110 and 120, functions were performed for the housing reports which paralleled those in Phases 114 and 119 for the population reports. (See above.)

#### **Subfile Creation--Phase 123**

In Phase 123 eight subfiles were split from the final sample detail file, based on characteristics of the population, as follows: (1) 5-percent sample; (2) migrant workers and members of the Armed Forces; (3) highly skilled and educated persons; (4) Government workers; (5) races other than white; (6) group quarters; (7) farm residence; and (8) Spanish language, Spanish origin, and Puerto Rican stock.

#### **Tabulations for Puerto Rico, Outlying Areas, and Americans Abroad**

The computer processing of population and housing data for Puerto Rico and the outlying areas and population data for Americans overseas was accomplished in a separate series of programs for each of these major areas. Processing for Puerto Rico most nearly resembled that for the United States. This was made possible by the close relationship of the basic questionnaires for the two areas. Publications and user tape files also were similar. Stateside programs therefore were adjusted to accomplish the necessary tasks.

For the outlying areas, a number of the initial U.S. programs were adjusted to accommodate the data, but only one tabulation (similar in content to the U.S. fourth count) was made to satisfy publication requirements. No user files were prepared.

Data processing for the overseas population required a completely different system. Punched input was used rather than microfilm. Only limited data were tabulated for addition to the appropriate U.S. reports. The remaining data were prepared for publication in one report, series PC(2)-10A, Americans Living Abroad.

#### **Subject Reports**

There were 39 Population Volume II (Series PC (2)) and nine Housing Volume VII (Series HC (7)) subject reports prepared during the 1970 census period. Each report concentrated on a particular subject, with detailed in-

### Publication Processing

formation and cross-relationships generally provided on a national and regional level. In some reports data for States and other smaller areas were also shown. Among the characteristics covered in the population subject reports were national origin and race, fertility, families, migration, education, unemployment, occupation, industry, and income. Housing reports included housing characteristics by household composition, and housing of minority groups, senior citizens, and mobile-home households. For titles and publication information, see chapter 10.

A special system was designed for processing the subject reports. This system consisted of three main operations--tallying counts, summarizing data, and preparing tables. In each operation, base or framework programs were created that were generally applicable to all of the reports. Standardized support programs were also developed that could be used to generate test decks and to display data. The key to the tally program was its ability to perform mass tabulations. In one report, for example, 1.5 million base cells were accumulated for each unit of processing. Normally, the tally data were derived from the weighted detail file or from one of the subfiles. Most of these files were maintained at the quasi-State level; in the summary process, the quasi-State data were consolidated into national, regional, divisional, and State levels. In the table-preparation phase, the summarized tally data were manipulated to meet the requirements of the publication. A framework table-preparation program, capable of referencing data in a standard pattern, was developed for this purpose.

The key to the programs for the subject report series was matrix logic. The table outline specifications were interpreted as a series of tally matrices; these matrices were frequencies, aggregates, and sum-of-square aggregates, depending on the qualities of the particular tables. The tally data were cross-tabulations of the data, often with overcategorization to aid in verifying the results. To assist in the verification process, a display was developed to present only the totals of each component within a matrix. An alphabetical index of components was created so that results could be compared.

In addition to the published reports, subject data were also made available on tape for census users. Generally, the user tapes contained more detail for each geographic area and for smaller areas than found in the printed reports. (See chapter 10 for details.)

Some of the physical characteristics of the subject reports were as follows:

	<u>Range</u>	<u>Average</u>
Tally matrices	6 to 137	53
Tally cells	13,838 to 1,373,345	357,600
Tables	2 to 372	42
Published pages	119 to 1,132	410
Programmer days	34 to 328	120
Computer hours (Model 1108)	2 to 32	15

During the census period a Bureau staff member, assisted by members of the Linotron coding unit, exercised a coordinating and guiding role in the editorial aspects of the 1970 census publications. A close relationship was maintained with the subject-matter divisions to ensure that their computer program plans were ready when required for execution; these plans were reviewed from editorial and systems viewpoints. A close relationship also was maintained with the staffs which prepared the data and systems programs, in order to be aware of these programs' capabilities, the techniques employed, and the problems involved. Advice was given on publication practices and standards, and assistance was made available for resolving problems. Liaison was developed with the publications planning staff, and each markup of copy was reviewed; however, as the end of the census period approached and the workload increased, only a cursory review of the subject reports (see above) could be performed.

The Linotron coding unit was responsible for the coding of all Linotron photocomposer production formats for tabular presentation of most of the population and housing census reports. This group was directly responsible for managing and using the Linotron computer information storage and retrieval system developed for the 1970 census statistical tables. (See chapter 10 for details.) At the peak of production, this group consisted of approximately 15 persons who interpreted and coded for Linotron the editorially marked-up tables for approximately 133,000 pages in 10 different publication series and reviewed the page-image listings for the base files as well as for the expanded files. In addition, they also did the expansion planning and markup for the 241 reports in the PHC(1) series (22,000 pages).

The work schedule during the peak periods usually consisted of both extended work days and work weeks. Production was affected by the need to hire and train new personnel, to resolve "markup" inconsistencies, to monitor the quality and quantity of encoding or punching of coded data, and to cope with recurring problems of lost or misidentified data tapes, as well as system deficiencies. The average subject report was produced in about 50 working days, depending on computer priorities and turnaround time and the timing and completeness of the printing. About 25 days were needed for the shortest report, while approximately 100 working days were required in one situation.

### Computer Time

The term "chargeable hour" used below is a unit of measure used at the Bureau's installation in evaluating the usage and availability of computer capacity to avoid the measurement problems inherent in the use of computers with varying central processing speeds. Chargeable hours are defined as the number of wall clock hours that a computer is available for use. Estimates of these hours are derived by applying average conversion factors to the recorded central processing hours delivered by the different computers (e.g., 1106, 1107, 1108) at the Bureau's installation.

<u>Fiscal year</u>	<u>Total chargeable hours used at Census Bureau installation</u>	<u>Number of chargeable hours used for 1970 census</u>	<u>Percent of total</u>
Total	<u>166,607</u>	<u>27,878</u>	17
1970	31,273	875	3
1971	36,349	8,173	22
1972	43,633	9,979	23
1973	55,352	8,851	16

The table above describes the number of hours chargeable to post-enumeration data processing, including the preparation of master summary tapes, publications, and the public-use sample. Excluded are such items as preparation of the ACG's (address coding guides), DIME (Dual Independent Map Encoding) files, and address registers, as well as the computer work involved in reports on housing quality and the Surveys of Components of Inventory Change and Residential Finance. The

1970 census computer time is shown as a percentage of the total chargeable hours on the Bureau's computers.

## COSTS

The data-processing costs for the Nineteenth Decennial Census shown in table 2 include depreciation, but they do not include the cost of general administration, other general expense, or capital outlay which were recorded only at the appropriation level. These costs are shown in the 1970 Census of Population and Housing cost summary in chapter 1 of this history.

The data-processing costs for coverage improvement and verification programs, such as the investigation and adjustment of field undercounts, the resampling of selected enumeration districts, the post-enumeration Post Office check, the "Were You Counted?" campaign, and the missed persons search were included in the costs for data collection (see chapter 5), and therefore do not appear in the table.

Table 2. Data Processing Costs by Fiscal Year

(In thousands of dollars, figures rounded)

Project title	Total	1966	1967	1968	1969	1970	1971	1972	1973
Total.....	29,545	1	27	257	629	2,194	20,128	4,370	1,939
Process planning.....	1	1							
General processing procedures, direction and control....	667		27	99	56	65	86	133	201
Evaluation of coding operation.....	20							20	
Preparation of breaker sheets, 100-percent and sample...	207			5	1	48	153		
Receipt, check in, and control, 100-percent.....	1,045			5	2	148	890		
Microfilming, 100-percent.....	1,734			1	11	483	1,238	1	
FOSDIC scan (Phase 1), 100-percent.....	818			7	8	69	734		
Review computer diaries, 100-percent.....	824			6	9	57	752		
Preliminary area sort (Phase 2), 100-percent.....	309			2	5	12	290		
First count (Phase 3), 100-percent.....	215				25	16	158		
Geographic identification A (Phase 4), 100-percent.....	87			5	5	3	74		
Geographic identification B (Phase 5), 100-percent.....	44			2	6	3	33		
First count tabulations (Phase 6), 100-percent.....	201			4	28	31	138		
Final area sort (Phase 7), 100-percent.....	82			2	1	1	78		
Second count (Phase 8), 100-percent.....	195			10	34	10	138	3	
Third count (Phase 9), 100-percent.....	245			2	31	30	44	138	
Review block computer diaries.....	248				15	16	204	13	
Tabulation result work, 100-percent.....	207				18	52	73	64	
Summarization of Second Count PB series (Phase 10), 100-percent.....	204				29	48	82	45	
Tract summarization (Phase 11), 100-percent.....	75				16	13	19	27	
Housing Vol. I summarization (Phase 12), 100-percent....	76				9	19	39	9	
Copy 100-percent microfilm.....	219						2	89	128
Mechanized quality control for coding operations.....	429					78	365	(11)	(3)
Control of sample questionnaires.....	661					9	616	36	
Sample selection for quality control.....	122				1	20	101		
General coding.....	3,344			8	25	210	3,097	4	
Place-of-work coding.....	3,214			6	16	141	3,020	31	
Industry & occupation coding.....	4,073			5	15	85	3,960	8	
Microfilming sample.....	1,568				7	22	1,327	212	
FOSDIC scan sample (Phase 100).....	731			1	4	3	426	297	
Sample computer edit (Phase 101).....	618			61	96	30	240	191	
Review sample computer diaries.....	738				14	5	366	353	
Weighting sample.....	118		1	18	17	17	16	66	
Area sort (Phase 102).....	88				2		47	39	
Geographic identification merge (Phase 103).....	163				25	12	23	103	
Place-of-work match (Phase 105).....	244				7	29	88	120	
Small-area tabulations:									
Population (Phase 106A).....	276				38	35	23	180	
Housing (Phase 106B).....	89				36	13	11	29	
Large-area tabulations (Phases 107 and 108).....	371				5	51	73	205	37
Large-area PD series, States and U.S.....	271					16	78	121	56
Large-area tabulation, Housing (Phases 109 and 110)....	131					37	25	69	
Large-area summary, Housing Vol. II.....	67					10	30	27	
Small-area PC series, States and U.S.....	240				13	63	25	139	
Tabulation result work, sample.....	165				2	1	19	33	110
Small-area tabulations, tract summarization.....	119				2	20	12	85	
Housing tabulations, small areas, States and U.S.....	209				3	40	40	126	
Population subject reports.....	1,290						17	304	969
Housing subject reports.....	287							22	265
Data processing, Housing Vol. VI.....	170							12	158
Review of local-area counts.....	217							199	18
Population and Housing management studies.....	23						23		
Quality control recordkeeping.....	148						118		
Data processing, miscellaneous.....	1,038					49	229	760	
ZIP code tabulations, 100 percent.....	108					10	32	66	
Supplemental forms.....	492					64	426	2	

## BIBLIOGRAPHY

- American Society for Quality Control. Administrative Applications Division Conference Transactions, Arlington, Va., March 12-13, 1970. Milwaukee, Wis., 1970. Pagination by sections. Includes the following sections:
- 3-G "Comments on Quality Control and Research in Data Processing," by George Minton. 23 pp.
  - 3-I "1970 Censuses of Population and Housing, Quality Control Applied to Automated Equipment," by Maxwell D. Jeane. 13 pp.
  - 3-L "Quality Control Applied to Clerical Coding," by Jean Biggs. 5 pp.
- Dolleck, Sol; Maxwell D. Jeane; and Francis J. Boucher. "Comments on the Planning, Organization, and Monitoring of the 1970 Decennial Census Processing System." Paper prepared for the Conference of European Statisticians, Washington, D.C., May 15-24, 1968. 15 pp.
- Drury, Robert F. "Electronic Computer Management in the U.S. Bureau of the Census." Paper presented at the Inter-American Statistical Institute Meeting, Caracas, Venezuela, October 16, 1967. 10 pp.
- Gorinson, Morris. "Integration of Manual and Computer Techniques for Editing and Coding and Pre-Publication Review of Tables." Paper prepared for the Conference of European Statisticians, Washington, D.C., May 15-24, 1968. 10 pp.
- Hornseth, Richard A., and Ralph E. Mullendore. "The Computers and U.S. Bureau of the Census Publications." Paper prepared for the Conference of European Statisticians, Washington, D.C., May 15-24, 1968. 17 pp.
- Lawrence, Charles B., Jr. "Computer Experience at the Census." Datamation, Vol. 12, No. 7, July 1966, pp. 45-52.
- Minton, George. "Inspection and Correction of Error in Data Processing." Journal of the American Statistical Association, Vol. 64, No. 328, December 1969, pp. 1256-1275.
- . "Some Decision Rules for Administrative Applications of Quality Control." Journal of Quality Technology, Vol. 2, No. 2, April 1970, pp. 86-98.
- O'Reagan, Robert T. "Relative Costs of Computerized Error Inspection Plans." Journal of the American Statistical Association, Vol. 64, No. 328, December 1969, pp. 1245-1255.
- United Nations. Statistical Commission and Economic Commission for Europe. Conference of European Statisticians. Automated Preparation of Input to Computers. New York: United Nations, 1965. 89 pp.
- U.S. Bureau of the Census. "Electronic Data Processing at the Bureau of the Census." Revised February 1965. Washington, D.C., n.d. 23 pp.
- . "Electronic Data Processing Equipment for Census '70." Washington, D.C., n.d. 9 pp.
- Voight, Robert B. "Relationship between Enumeration and Processing Plans for the United States 1970 Census of Population and Housing." Paper prepared for the Conference of European Statisticians, Washington, D.C., May 15-24, 1968. 16 pp.
- Volk, Masey. "Automated Input for the 1970 Census." Journal of Micrographics, Vol. 3, No. 2, Winter 1969-70, pp. 57-62.
- Volk, Masey; Anthony Berlinsky; and Sol Dolleck. "U.S. Census Bureau: 1790 Tallies to 1970 Microfilm Scanning." Paper presented at the 19th Annual Convention of the National Microfilm Association, San Francisco, Calif., April 28-May 1, 1970. 7 pp.