PROCEDURES FOR THE 1960 CENSUSES OF POPULATION AND HOUSING

By: Morris H. Hansen, Assistant Director for Statistical Standards, Bureau of the Census
U. S. Department of Commerce

Major changes are being made in the methods for taking the 1960 Censuses of Population and Housing as compared with previous censuses.[1] Some are strictly procedural and should have little or no effect on the nature of the results to be published. They will reduce the cost or speed up publication of results of the Censuses, and may be of interest to statisticians on this account as well as because of their potential application in other areas of statistical collection and compilation. Other changes in methods should improve the accuracy of results published. The changes in procedure represent an evolution from previous census methods, and the new aspects have been evaluated in various ways including tests, before, during, and subsequent to the 1950 Censuses. We have had experience in other recent censuses with some of the new procedures. The plans call for decisions and detailed planning much earlier than has been the case in the past.

Among the principal changes in methods now planned for the 1960 Censuses of Population and Housing are the following:

(1) Extension of sampling. Sampling will have a fuller role in the 1960 than in the 1950 Censuses. The complete population census will be limited to a few basic items. Information on other subjects will be collected from a 25 percent sample of households whereas in 1950 some of the subjects were handled by 100 percent enumeration and others by 20 percent and 3-1/3 percent samples. Sampling will serve a similar role in the housing census.

(2) New Equipment. All data to be tabulated in the census will be recorded on specially prepared forms by means of positioned marks (check boxes). These forms will be microfilmed, and newly developed electronic equipment will convert the data from microfilm to magnetic tape without manual punching of cards. The magnetic tape will serve as input and the returns will be edited and tabulated in the computer. The computer editing can be so controlled that it will automatically dispose of most inconsistencies and non-response but will print out for manual review those which might seriously distort subsequent tabulations. Much of the final results will be published by direct offset reproduction of the high speed printer output.

(3) Enumeration methods. Although enumerators will visit every home, much of the information will be furnished by forms which the householders themselves have filled out. There will be a two-stage operation in most areas. The 100 percent questions will be distributed in advance through the Post Office, with a request for respondent cooperation. The questionnaire, if filled out, is to be held for review and completion by the first stage enumerator. The sample questionnaire will, in a limited number of areas of low density, be filled out at the first stage, but in most areas will be left by the original enumerators at 25 percent of the households, with a request that it be filled out and mailed in. Selected enumerators who have worked on the first stage will be trained for the second-stage operation and will review and transcribe the returns to the final census schedules, and revisit the sample households as needed to complete the information. This procedure provides an opportunity for participation in the responses by all members of the household, and reduces the contribution of enumerators to response errors.

The objectives of these procedural developments are threefold. First, we expect to make substantial reductions in the time it takes to publish results; second, we hope to improve the quality of the censuses; and third, we expect to achieve the first two objectives at less cost than the 1950 cost levels adjusted for the increase in population and salary and price changes.

Extension of Use of Sampling in 1960 Censuses

In the 1950 Population Census, a 20 percent sample of persons was used for the collection of data on education, income, migration, and other subjects, and a 3-1/3 percent sample was used for fertility questions and certain other topics. For 1960 we shall, in addition, transfer citizenship, place of birth, employment status, occupation, industry, and related questions to the sample and increase the sampling fraction from 20 to 25 percent. The sample in 1960 is a sample of households. Remaining on the 100 percent population census will be the listing of the population, relationship to head, age, sex, race, and marital status. The 100 percent questionnaire will contain no questions requiring manual coding, and the questionnaires can go directly into microfilm and through processing without any delay for manual operations. Thus, a definite principle in deciding which questions to include on the sample was whether manual coding of the returns is necessary before processing.

Of course, the other and fundamental principle of deciding whether questions should be collected for 100 percent of the population or on the 25 percent sample is the requirement for data. Most of the items chosen for 100 percent coverage are basic demographic characteristics on which relatively precise statistics are needed for small as well as large areas, and for which past experience shows that relatively high accuracy can be achieved in the census. Many of the items on the sample involve concepts that are relatively difficult to measure. The use of sampling permits coverage of more items in the census, faster compilation, and lower costs.

Sample selection will be accomplished by designating every fourth household in order of visitation by the enumerator for inclusion in the sample. All members of each selected household will be included, as will data for the housing unit. Somewhat different procedures will be used for selecting the sample within institutions, large lodging houses, and other special dwelling places. In 1950 a sample of lines was predesignated on the population schedules, and the sample data were obtained from the persons enumerated on these lines.

The choice of housing items for complete coverage or sampling will follow principles similar to those described for population. However, for cities of 50 thousand or more population we plan to publish summary housing characteristics for individual blocks, and the items tabulated for blocks will be on the complete census. The items collected for all housing units in these cities will include, tenure, rent or value, number of rooms, condition, plumbing facilities and certain other questions. In the remainder of the country tenure and occupancy status, number of rooms, type of housing, condition, and plumbing will be collected 100 percent, and the remainder of the questions will be collected on a sample. For housing the sample will be somewhat more complex, however, in that some of the sample items will be collected for a 25 percent sample, and others will be from 20 and from 5 percent samples.

In both population and housing the items collected from the 25 percent (and also the housing 20 percent) samples will be tabulated in substantially as great detail as would be the case if the items were collected 100 percent. For most large area statistics, such as for states, metropolitan areas, and large cities, the use of this sample rather than 100 percent enumeration should have a barely perceptible effect in decreasing reliability of results. For small area statistics, such as census tracts, and very detailed tabulation cells for large areas, the sampling errors will be relatively large, but even here the accuracy for some items will not be substantially weakened as compared with 100 percent coverage, primarily because response errors for these items are relatively large.

We had hoped that all of the economies from the sampling would be available to finance improved quality and coverage. While some of these gains will be achieved, most of the gains from sampling will probably take the form of reductions in the total cost of the census.

A major gain from the extension of sampling is the ability to publish results much sooner (this is a joint gain from the extension of sampling and the application of electronic equipment, together with a great improvement in the advance planning of the census).

Electronic Equipment

Computers. The first Univac was delivered to the Census in 1951 and was used for compiling some of the 1950 Census. While it contributed only a small amount to that Census, the experience taught us much. During the next five years we put it into use on our current work and on many special and service projects. We needed additional capacity for the compilation of the 1954 Censuses of Business and Manufactures, and jointly with the Internal Revenue Service acquired a second Univac in January 1955.

About the time the processing of the Economic Censuses was virtually completed in 1956 we addressed our attention to the question of the computer capacity we would need to process the work load we could anticipate during the 1959 to 1962 period.

When we matched our anticipated work load against the production capacity of the Univac I's, we decided that we would need a minimum of eight Univac I's in full-time operation during the peak of our requirements. Since more advanced computers were becoming available, we embarked on a study of the comparative merits for our work of several of the new systems. The study was completed early in 1957 and resulted in a decision to replace our two Univac I's with two Univac 1105's. The 1105 system is a modification and improvement of a computer system known as the 1103A, designed primarily as a scientific computer. The modifications, which were introduced to meet the Census requirements, adapt the equipment for exceedingly effective large scale data processing work as well as for scientific computing.

The first 1105 was installed last October and the second one in December. Until about June of 1960, the 1105's will be used for current work and for processing the 1958 Censuses of Manufactures and of Retail and Wholesale Trade, and Service Establishments. Thus, we shall have considerable experience on our new computers before embarking on the tremendously large scale work of the 1960 Censuses of Population and Housing.

The two 1105 systems installed at the Bureau of the Census headquarters will be capable of delivering about 3,000 hours per quarter of productive time by operating seven days a week 24 hours a day. Approximately twice this capacity will be needed during our 1960 Census operations. To provide the additional capacity the Bureau of the Census is cooperating with the University of North Carolina and with the Armour Research Foundation of the Illinois Institute of Technology. In each of these universities an 1105 system compatible in all respects with the Census systems will be installed. In return for what will be effectively a prepayment for time on its computer, each university has agreed to make about two-thirds of the productive time on its computer available for Census use during the period of peak work load. In addition, the Census will have access to a smaller amount of time at each university for a period following the completion of the processing of the 1960 Decennial Censuses of Population and Housing.

Printers. One of our early lessons with electronic equipment was that the success of a large scale electronic data processing installation depends as much on the reliability and performance of its off-line auxilliary equipment as it does

on the central computer. A reliable high speed printer is one of the important auxilliary units, and we have two high speed printers in operation. We hope to program our output from the computers in such form that the output from the printer will be ready for final review and photographic reproduction. The ability to accomplish this through the flexibility of a computer system and a high speed printer is considerably greater than with punched card equipment. This utilization of the computer and printer will aid in speeding up and in lowering costs in the census production. We plan to do direct reproduction of printer output for the tables that are repetitive in general format for many areas or other classifications. Some of the tables, however, will be put together most effectively by manual methods, as in the past, but even in these cases the computer and printer system should simplify the manual operation to a considerable extent.

Fosdic. One additional auxilliary electronic device is particularly essential to our plans for processing the 1960 Censuses of Population and Housing. This is equipment developed for us by engineers at the National Bureau of Standards. It is known as FOSDIC, a name consisting of the initial letters of Film Optical Sensing Device for Input to Computers.

It was about 14 months after the enumeration date of April 1, 1950, that the last card was punched for a person enumerated in the 1950 Census. To improve this time schedule significantly we could plan to increase the staff of key punchers or we could search for some automatic method for accomplishing the translation from Census schedules to a tabulation medium. We described the problem to personnel of NBS and after several months of study they proposed FOSDIC.

The FOSDIC forms are designed so that answers are recorded as marks in specific locations (similar to the way data are recorded in cards by punching holes in specific locations). Almost any writing instrument can be used. The completed questionnaires are first photographed on 16 millimeter microfilm, and FOSDIC scans each frame of microfilm at a rate of about 100 to 200 a minute, depending on the amount of information recorded. It automatically reads recodes, and records on magnetic tape the intelligence that has been entered on the schedule, FOSDIC can be programmed to scan the data recorded in a microfilm frame in almost any desired sequence and has a great deal of flexibility for adjusting itself to various sizes and designs of questionnaires and for tilt, shrinkage, or expansion of the copy when it is filmed. FOSDIC has the further useful ability to identify the dominant or heaviest mark in a particular area, thereby making it possible to correct an answer by normal erasure. Experimental work with a completed prototype indicates that its reliability as an input preparation device will result in greater accuracy than has been achieved in the past with either mark sense punching or manual key punching. One of the four production models of Fosdic being built at Census has been completed and is now undergoing final checking and adjustment.

Population 100 percent Tabulations. We contemplate the use of two basic types of FOSDIC census schedules, the 100 percent (short) schedule for both population and housing, and a longer sample schedule. The basic information recorded on the 100 percent schedule will also appear on the sample schedule for these households. These two forms are used to help speed up compilation and publication. We expect to reduce the time for publication of census results by an average of about a year and a half as compared with the 1950 Census, and to compile and publish the census results obtained from all persons (the 100 percent information) within a few months after the returns are received.

It is important to recogrize that the information recorded on the 100 percent Fosdic schedules will not require coding in the office. This is in contradistinction to data on such characteristics as state or country of birth, occupation, industry, etc., which will appear on the sample schedules and which will be converted manually to number codes in the processing office before they are tabulated.

The legislation under which we conduct the Census requires us to report the total population not later than eight months after the beginning of the field enumeration. To meet this requirement we must determine only the number of persons in each state; we do not have to classify them by their age, sex, occupation, or any other characteristic. This initial or "apportionment" report has always been the result of counting by hand—no non-manual method for obtaining this count on a satisfactory time schedule has heretofore existed. However, for the 1960 Censuses our plans comtemplate that we will microfilm the 100 percent schedules, process the microfilm through FOSDIC and the resulting magnetic tape through our computers by the end of October 1960. Thus, we expect to have not only the population count from the machines, but also the tabulations of 100 percent data on the time schedule on which the hand count alone was completed in 1950.

Without electronic data processing equipment we do not believe we could work to such a tight time schedule. However, the computers alone do not bring our hopes within range of accomplishment. Only with the elimination of manual coding required for the complete census schedules, the ability rapidly to prepare our input medium which FOSDIC provides, and use of the computers does our objective become feasible. In addition, this time schedule requires advance planning beyond any level accomplished in earlier censuses, and it calls for doing the re-allocation and searching for duplicates of persons enumerated away from home as a part of the field collection operation instead of in the processing offices as has been done formerly.

The time schedule allows about four months over-all after completion of the field work for microfilming, preparing the tapes on FOSDIC, and putting them through the computers, and only about three months for any one of these operations. Three months to tabulate data for 180,000,000 persons is indeed short. Only with the most detailed

kind of advance planning can we hope for success. Our plans for this part of the work can be described as having two major characteristics. First we are organizing the job in such a manner as to reduce the number of work units to be separately identified and controlled from 180,000,000 to about 260,000 (the approximate number of enumeration districts into which the country will be subdivided). Second we expect to have "dry runs" of processing the Census between now and the time we actually begin processing the returns.

With respect to our plans for reducing the number of work units from 180,000,000 to 260,000, it was mentioned earlier that the microfilms of 100 percent content will be photographs of the schedules as prepared by our enumerators with no manual coding in the office. The only detailed pre-tabulation inspection of these schedules will be by field crew leaders and will for the most part occur during the course of the enumeration.

We recognize that at the time the schedules are microfilmed not every question will be adequately answered for every person enumerated. There will be some omissions and obvious inconsistencies. The detection and correction of such errors is what is commonly called editing, and we will direct the computer to perform the essential edits. In our judgment, this is, in many ways, a better way to edit than by manual inspection. The computers will do the work much faster. They will apply the editing rules individually to each of the 180,000,000 people, they will do it absolutely consistently, and they can be directed to keep track of how often they apply each rule. An illustration here may help to illuminate this point. The following simplified rules might be incorporated in the 1105 instruction program which would be so arranged that they were applied only when the question on the marital status was not answered.

1. If the person is under 15 years old—proceed to the next person (we do not tabulate marital status for children)
2. For a person 15 years old or older examine relationship
   a. for wife or for a male head of household followed by a wife assume married
   b. for other relationship, assign marital status by using the answer for the preceding person in this class of the same sex-age group for whom marital status was given and in addition—this is important—add "one" to a tally of the number of times an entry for marital status was imputed by the application of this rule.

In this illustration we see how the computer might proceed with the preparation of a tabulation by supplying missing information. This will seriously distort the resulting tables only if it is done too often, and the computer is keeping track for us of how often it is done for each work unit. The final editing rules may be quite different from the simple illustration above but the principle will be the same.

The technique for combining editing with tabulation leads to the plan for making 260,000 the effective number of work units to be controlled rather than 180,000,000. We expect to tabulate the 100 percent population statistics for each enumeration district (E.D.) separately. As each E.D. is completed our instructions to the computers will direct them to evaluate the resulting statistics and determine whether they are acceptable or should be subjected to further review and perhaps correction. Thus, we can instruct the computer to accept an E.D. only if there is a "reasonably small" difference between the field count of population and dwellings and the computer count, and also if the number of computer-kept tallies of the number of imputations it has made for each characteristic is "tolerable." We will have complete control of the definition of "reasonable" and "tolerable." The computers will rigorously and unerringly apply whatever standards we establish. We expect that the standards finally used will be such that about 90 percent of the ED's will be acceptable at the first run through the computers and that the remainder numbering about 25,000 ED's will be a manageable number of units which will be identified by the computer as unacceptable, and require manual investigation, correction, and rerunning.

The art of planning and computer programming is extremely intricate and requires the exercise of planning abilities of a high order and painstaking attention to detail. Also, if we are to achieve the desired time schedules, it requires that complete and final tabulation plans must be made available to technical programmers at a much earlier date than was required by the processing system previously used. With such timing the possibilities for flexibility and for improvisation of tabulation plans as you go along are extremely limited. We have for the past several months been writing, rewriting and testing some of the programs to be used on 1960 Census data during the summer and fall of 1960 and throughout 1961. This process will continue for more than another year before we are finally satisfied that we are ready for the big job.

We are certainly not sanguine about our ability to implement these plans without encountering serious difficulties. Here our plans to spend most of the next year on "dry runs" are important. During late February and March we plan to enumerate a pretest of the 1960 Census of Population. The areas we have selected for this dress rehearsal contain about 100,000 inhabitants and will allow us to anticipate many of the problems we will face in 1960. Also, we may run one more test in the summer of 1959 after the final census schedules are released for print. We plan to microfilm, process through FOSDIC, and tabulate the schedules resulting from these tests as many times as necessary between now and the summer of 1960 to acquire real skill in, and familiarity with the whole sequence of procedures involved.

Housing 100 percent tabulations. To this point our remarks concerning our electronic data processing equipment have been oriented to the 100 percent population tabulations. Our comments apply equally to the 100 percent housing tabulations except that for the Housing Census there is no such

early time requirement imposed by law. Nevertheless we hope to process the 100 percent housing data on approximately the same time schedule we have set for the 100 percent population data.

Sample data. We plan to use the same complex of microfilm, FOSDIC, and 1105 computers to process the sample data for population and housing. For the sample population schedules the manual coding of occupation, industry, place of birth, and other questions must precede microfilming. While we are processing the 100 percent schedules this coding will be started. The microfilming of the sample schedules will begin as soon as the microfilming of the 100 percent schedules is completed.

The volume of microfilm work and FOSDIC operation for the sample schedules will be greater than the work load for the 100 percent cases. The workload on the computers for the sample will be several times that for the 100 percent tabulations when measured in terms of computer time required. Whereas we hope to edit and tabulate the 100 percent data with one pass of the magnetic tape through the computer (except for rejected ED's and for summarization runs) there will be several passes necessary for the more detailed sample tabulations.

Our electronic data processing equipment has already demonstrated its utility for applying a variety of estimation formulas to sample statistics. The sample design for the 1960 Censuses is a fairly simple one. Nevertheless, the electronic processing will aid materially in obtaining estimates from the sample. We expect to use a ratio estimate, separately by age-sex groups and perhaps by size of household groups, still to be specified in detail. The ratio estimates will be applied by small areas, and fixed weights determined for each person will be carried through subsequent tabulations. Such ratio estimation would not have been practicable without the use of electronic computers. The ratio estimates will reduce the sampling variances and result in close agreement between sample and complete census tabulations.

Census Enumeration Methods

In much of the country, perhaps 85 to 90 percent, we expect to divide the collection of census data into two stages. The first stage will be the complete census canvass. In advance of this canvass, questionnaires containing the 100 percent questions will be distributed through the Post Office to give respondents an opportunity, through self-enumeration, to prepare considered answers before the enumerator calls. There will be supporting publicity. The first-stage enumerators will canvass all households in their assigned enumeration districts, as in previous censuses. They will transcribe to the FOSDIC forms from the questionnaires prepared in advance by respondents, or, as necessary, will ascertain by interview and record the information on FOSDIC schedules. In this first-stage enumeration every 4th household will be designated for the sample, and the sample questionnaires will be left with these households, to be filled out by the respondents and mailed in.

Second-stage enumerators will be selected from among the first-stage enumerators about April 15, after most of the initial canvass is completed. They will be trained on the sample questions, will review and transcribe the mailed returns to FOSDIC forms, and follow up by telephone and personal visitation as necessary to complete the sample information. In each stage the necessity of transcription by the enumerator calls to his attention any questions for which information has not been provided in the self-enumeration form, and provides an opportunity for examination of each response for acceptability. He interviews and collects the data not entered on the self-enumeration form.

In about 10 to 15 percent of the country, especially in the more sparsely populated areas, a one-visit procedure will be used, with both 100 percent and sample questions covered by the same enumerator on a single visit. In the areas where the one-visit procedure is followed, as elsewhere, the advance self-enumeration form will be distributed through the Post Office for respondents to fill in. This form will contain only the questions on coverage and content to be covered in all households. On occasion, in these areas, sample forms may be left to be mailed in, to reduce callbacks.

It has long been recognized that the census procedures place a particularly heavy burden on enumerators. They must absorb information concerning techniques of enumeration, the use of maps, administrative procedures, identification of separate living quarters, identification of residents and nonresidents and determination of who should be enumerated at a particular location, and designation of sample households. In addition, they must learn to apply the many concepts and definitions relating to the various subjects in the 100 percent and sample inquiries. Our field staff members believe that by separating the work into two stages, one primarily a canvass for census coverage and a few basic questions, and the other for the more difficult questions, we can more effectively train the enumerators for each stage separately, and can control their work more adequately, than under the one-visit approach. Also, the initial census canvass can be completed more rapidly, and thus reduce the effect of population movement on census coverage.

Under the two-visit procedure we propose to train about 160,000 enumerators on the first stage of the work and as they complete their first work stage, to choose about one-third of them to be trained for and to carry out the second stage operation.

The introduction of self-enumeration represents the first time this method has been widely used in the U. S. Censuses of Population and Housing. Experimental tests have been run on the use of self-enumeration at various times, including tests in 1948, and earlier, a test as a part of the 1950 Censuses, and tests in 1958. These tests have demonstrated widespread public cooperation, and show that the joint method of self-enumeration with enumerative follow-up is a feasible operation, and can be accomplished at about the same

cost level as with a regular direct enumeration census. Actual cost levels will depend upon the effectiveness of public cooperation.

Our plans do not assume that we shall receive substantially complete public cooperation, but we are optimistic that we shall receive very favorable public response in a national census undertaking. In the test conducted in a few counties in the 1950 Census the questionnaires were to be mailed in and about 95 percent were received through the mail. In this case an average of more than 80 percent of the individual responses on questions were acceptable and did not need further inquiry, but some questions were unacceptable on many returned forms, and follow-up on one or more items was required for a high proportion of the households. We hope to improve the self-enumeration forms and publicity and achieve even better cooperation in the 1960 Censuses.

The self-enumeration approach supplemented by follow-up enumeration as necessary is designed to take advantage of the joint contributions that can be made by respondents and enumerators. Our tests and research show that in a large army of more than 150,000 enumerators, recruited and trained in a short period, as they must be for the census, a considerable number will sometimes misinterpret the census instructions to enumerators. Particularly serious is the fact that an interpretation on the part of an enumerator may affect the results for the entire area that he covers. We have learned that such variations in interpretations consistent within the work of an enumerator but varying among enumerators, can significantly affect census statistics, especially for small areas and small cells. Self-enumeration reduces the effects of such tendencies toward consistent misinterpretations. At the same time, this approach allows more mature consideration of the responses to the questions and permits the best informed respondents to participate.

These comments on the effectiveness of self-enumeration have been directed particularly at the quality of information collected on various subjects. We hope, also, that the self-enumeration forms and the two-visit procedure will aid in improving coverage of the census. The initial short (100 percent) self-enumeration form provides for listing all persons who are members of and live in each household, whether present or not at the census date, and also all persons present in the household, whether or not they live there.

This approach, we believe, should help improve the coverage of less closely attached household members, (related members other than the wife and children of the head, and especially lodgers, guests, and persons with no usual place of residence). These groups have proved particularly difficult to cover adequately in a census. Perhaps providing for everyone on the self-enumeration form will result in a more complete listing of persons who should be listed as living in a household. This procedure should, in addition, improve our ability to pick up and reallocate people in private households who live elsewhere but have no one there to report for them.

We had hoped to arrange for the Post Office to check our census coverage, but this may not be feasible with the funds to be made available. This check would have taken the form of furnishing a card to the Post Office for each address covered in the Census. The mail carrier would sort these cards as though they were to be delivered to the indicated addresses, and then identify any addresses or households known to the postal carrier but missing from the census. Tests have shown that this would be an effective device for improving census coverage. It now appears that our resources may not permit this and certain other steps designed to improve coverage.

We do expect, nevertheless, to introduce improved quality control operations in the field collection. We have not been able previously and we do not expect to be able to impose a rigorous quality standard and then inspect, identify, and reject unsatisfactory work sufficient to insure conformance with the specified standard. But we do believe we can provide enough sample inspection to identify the weaker enumerators and the weaker spots in the census, and provide for their improvement. The difficulty, of course, is that most of the census canvass is completed in a couple of weeks, and adequate inspection in a few days covering a sample of all enumerators' work would require very substantial resources. But we do believe that real progress can be made in inspection and improvement with the limited inspection and controls that can be established with available resources. It is in this respect that a Post Office check, were it feasible, would have been especially desirable because it would make effective use of a going organization that already has an extensive acquaintance with the people and where they live.

In summary, while we shall not be able to accomplish all of the improvements in method that we would regard as desirable, we are making many important advances in census methods. These should markedly increase the timeliness of census publications, and reduce costs, and we hope also will substantially improve quality. We shall undertake to evaluate the effectiveness of these and of certain alternative methods in special evaluation and experimental work to be conducted as a part of the census.

---

[1] A detailed description of the 1950 Census methods is given in The 1950 Censuses: How They Were Taken, Procedural Studies of the 1950 Censuses No. 2, Bureau of the Census.