David L. Kaplan, U.S. Bureau of the Census

The first reaction to the title of this paper might properly be "Why?" Do we really need a research program on data-collection techniques for the 1970 Census? After all, the results of the 1960 Census are accepted by the great mass of users as accurate and satisfactory for their purposes. The Census publications are neat and reasonably clear, the columns add up and the rows add across. But for an audience of statisticians, it is really unnecessary to explain that there were certain inadequacies and insufficiencies in the 1960 program which could be remedied at least partially by a more effective system of data collection.

Among the Decennial Census totems to which we statisticians at the Census Bureau pay constant homage are better coverage, improved quality of subject statistics, decreased cost, and speedier preparation of results. To help attain these and related objectives, we carry on a continuing research effort. The mail census program is part of this work.

As concerns coverage--that is, the count of people and their housing units--some techniques for reducing error appear feasible. For example, if we could compress the enumeration period into a shorter time, the likelihood of missing or double-enumerating people on the move could be minimized. Also, a gain could be made if the listing of dwellings was verified more thoroughly than the traditional approach of having the crew leader spot-check the enumerator's canvass. In this connection, a most useful contributor might be the letter carrier since he makes a virtual house-to-house tour every working day of the decade.

To improve the quality of subject statistics within the existing constraints of technical and financial resources, the best hope seems to be through the extension of the use of self-enumeration. Our experience in 1960 indicates that, on the whole, we get better data through self-enumeration than through enumerator interview. It is also true, of course, that the quality of subject statistics would be raised by improving coverage; significant distortions in certain subject distributions undoubtedly result from the differential underenumeration among various population groups and geographic areas.

To cut the cost of data collection, the most promising approaches would appear to be (1) a reduction in the house-to-house canvass by the enumerator and (2) the greater use of telephone to obtain missing information. On the former point, the mailman might well play an important role.

With regard to earlier completion of the results of the census, the data-collection system can contribute in two ways. One is by completing and sending in the census questionnaires in such manner that they can be processed most efficiently. The second way is quite obvious--the sooner the field work gets done, the sooner the processing can begin, and the sooner the results can be compiled.

Another problem which arose in 1960 and which is likely to plague us again in 1970, is the difficulty of recruiting and maintaining a temporary field staff to collect the information. Using the 1960 data-collection procedure again in 1970 would require approximately 200,000 field workers. Considering labor market conditions and the rates of pay which the census can offer, there is likely to be serious difficulty in staffing such a field organization, particularly in the big cities. To the extent that the required number of temporary field workers can be decreased, the problem would, of course, be diminished.

These and related considerations led us early in the 1960's to embark on a research program to test the expanded use of the mails for the 1970 Census. We had had considerable experience with the use of the mails in the 1960 Census and in certain pre-1960 tests. In 1960, an unaddressed Advance Census Form limited to the 100-percent subjects was delivered to all households across the United States by the Post Office. The householder was asked to fill out the form and hold it for the arrival of the enumerator. Also in 1960. for about four-fifths of the country, those households in the 25-percent sample were left with a sample questionnaire booklet which they were requested to fill out and mail back within three days. Our experience with this mail-back approach in 1960, the large-scale experiments conducted during the 1950 Census in the Columbus, Ohio, and Lansing, Michigan, areas, and a substantial test conducted in conjunction with a special census of Memphis city in 1958, all encouraged the belief that a mail-out/mail-back technique might be developed which would yield gains along some and perhaps all of the fronts previously mentioned.

Looked at operationally, there are three vital keys to the success of a mail census. One is the adequacy of the mailing list. The second is the public response which determines the amount and complexity of the follow-up work required. The third is whether the management of a mail census field office can be handled successfully by a temporary organization.

We started the systematic research program in August 1961 with a test in Fort Smith, Arkansas, a city of about 60,000 population. The enumerators made a house-to-house canvass and left a very short questionnaire for householders to mail back. Approximately 86 percent of the households mailed back the questionnaire. Also the list of addresses prepared by the enumerators was used to check an office-generated register of addresses which was based on our 1960 Census records updated through recent building permits and checked by the local Post Office. The results indicated that the corrected office-generated register was at least as complete as the enumerator listings.

In June 1962, we went back to Fort Smith and also went to Skokie, Illinois (another city of about 60,000 population) to test another variant of this approach. We had the Post Office deliver questionnaires addressed from an updated 1960 Census list. A coverage check showed the miss rate for housing units to be about one percent, which compared favorably with the 1960 Census national experience. In each city, close to three-quarters of the householders mailed back their questionnaires.

In April of 1963 we experimented again, this time in Huntington town, Long Island, New York, covering about 150,000 population. For most of the area, we based the mail-out on an updated 1960 Census address register, with a post office check for omissions at the time of mail delivery. In the remaining, essentially rural portion of the area, enumerators made a house-to-house canvass during which they left questionnaires to be mailed in. Also, for the first time in this series of tests, we used both long and short questionnaires. We thus simulated the 1960 (and anticipated 1970) approach whereby most of the population is requested to answer only a limited number of questions and a sample of the population is asked to answer the full range of subjects covered by the census. A coverage check revealed the miss rate for housing units in the address-register portion of the area to be about one percent. The mail return rate was 85 percent for both the short and long questionnaires, a very encouraging sign that the public was willing to cooperate even in answering an extensive questionnaire.

Our experience to date was, thus, quite encouraging on all fronts--public cooperation appeared to be on a satisfactory level, the potentiality for developing an effective mailing list was indicated, and the feasibility of the operation in the field seemed probable. We, therefore, embarked on a larger research program financed by a special appropriation for this purpose. It is noteworthy that officials of the Commerce Department, Budget Bureau, and the Congress recognized so early in the decade that preparatory work on the next Decennial Census should begin and that a feasibility testing program should be financed; more than one million dollars was spent on two large-scale experimental censuses in 1964 and 1965.

In May 1964, we took a special census of the Louisville, Kentucky-Indiana, Standard Metropolitan Statistical Area, covering close to 800,000 population. For mail-out purposes in the city portion of the area, we used the 1960 Census address list, updated with building permit and utility data. Outside the cities, we used a list prepared through a house-to-house canvass by Census personnel about a month prior to Census Day. At mail delivery time, the letter carriers informed the local census office of residential addresses for which they had no questionnaires; the missing addresses were added to the mailing list. Short and long questionnaires were sent out, essentially repeating the subject content of the 1960 Census. The questionnaire return rate was 88 percent, with only a few points difference between the short- and longform rates. An intensive sample check showed the coverage errors for both occupied housing units and population to be below the 1960 national levels. (The 1960 program did not yield coverage error estimates for individual areas.)

For the second large-scale feasibility test census, conducted in April 1965, we selected the city of Cleveland, Ohio. We chose Cleveland purposely to provide a rigorous test of the mail technique in a large cosmopolitan industrial city in which we had substantial enumeration problems in 1960 and in which we could expect similar problems in 1965 and in 1970.

Some exploration after the Louisville project suggested that we could do better on several accounts by using a commercial mailing list for the register of addresses instead of going through a trasnfer to tape and updating of the addresses on the 1960 Census schedules. While commercial lists do not contain the detailed geographic codes needed for census purposes, they can provide certain important economies and flexibilities over the 1960 materials. The procedure we used was as follows. In December 1964, the commercial list of approximately 300,000 addresses for Cleveland city was printed out on individual labels, each of which was affixed to a card. In January 1965, the cards were given to the local Post Office to check for omissions, revisions, and deletions. The corrections provided by the Post Office were made on the computer tape record. This was done in February, at which time we also inserted the necessary geographic codes, applied the required field control numbers, and designated each fourth household to receive the long-form questionnaire. By early March, individual address labels showing the necessary control designations were printed out and the labels affixed to the appropriate mailing pieces. Also printed out and bound in a separate book was the list of addresses in each assignment area. By March 20, all of the mailing pieces were in the hands of the Cleveland Post Office.

On Monday, March 29, the mailing pieces were delivered by the letter carriers. As they sorted and delivered the census mail, the letter carriers were instructed to make note of any housing unit for which they did not have a mailing piece. These missing addresses were transmitted as rapidly as possible to the local census office, where they were added to the address list for the particular area and mailing pieces sent out.

Householders were requested to send back their completed questionnaires on Census Day--Thursday, April 1. This request was contained in the covering letter on the questionnaire and was also the subject of a very intensive local publicity campaign. The local office checked the mail-back questionnaires for acceptability. Unacceptable returns were followed up by enumerators through telephone or, if necessary, by personal visit. Nonresponse cases, including vacant housing units, were followed up through personal visit by enumerators.

Eighty percent of the householders returned their questionnaires and, as in Louisville, there was very little difference in the return rate as between the short and long forms. Another encouraging aspect was that the great bulk of returns arrived on or shortly after Census Day (also much like the Louisville experience). By April 3, questionnaires had been received from about 70 percent of the households, which meant that the count of more than two-thirds of the population had been accomplished in just a few days, and the chances for error resulting from movement for these people had been minimized. The coverage experience in Cleveland city for occupied housing units and population was on a par with or better than the national averages in 1960 after standardization of the Cleveland results for the proportion of units and population in multiunit structures (where coverage is lower than in single-unit structures).

Our total experience in all of the research program to date has been gratifying on most all issues. At least in built-up areas which use citytype addresses, and perhaps even in rural areas, the mail approach appears to provide the various benefits we are looking for. There are still many problems to be resolved--for example, how to raise the acceptability rates for the longform questionnaires, what procedure to use in rural areas, and how to get better coverage in the congested areas of our big cities. We will be working on them intensively in the time remaining before the final 1970 Census materials are locked up. But the decision has been reached to use the mail approach for the larger part of the American population.

Note should be taken of at least two important associated advantages of the mail system. One, it provides greatly increased accuracy in the identification of city blocks, and also permits the identification of individual sides of blocks in the highly-detailed computerized geographic coding system which the Census Bureau is developing for the 1970 Census. Second, the address list, if kept updated after the 1970 Census, will be useful for later sample selection purposes, for the 1980 Census or the proposed 1975 middecade census, and for the preparation of local population estimates.

Our next large experimental census is planned for the metropolitan area of New Haven, Connecticut, in April 1967, where we will be refining our techniques and testing subject content. Over the next year, we will also conduct a number of smaller field tests for the same purposes. By the fall of 1967, we will be in earnest preparation for the "dress rehearsal" census scheduled for April 1968. In that project, we will no longer be trying out alternatives; the focus will be on the processes and materials we expect to use in the 1970 Census itself.