DISCUSSION

A. Ross Eckler, Consultant

Dr. Taeuber, the organizer of this session, is to be commended for planning a well balanced program covering an important, but often neglected stage in the production of statistics, a modern data delivery system, and the creation of a mechanism outside the government for utilizing more of the products of a national census. If I may shift the order to proceed from production through final use, Mr. Waksberg has reported on the plans for evaluating the 1970 Census of Population in the United States, work which I regard as the final stage of the production of a census. Dr. Murphy has told us about the means for the delivery of data in various forms from the 1971 Census of Canada. And, third, Dr. Lee has described a cooperative effort to establish a mechanism for the fuller utilization of material from the United States Bureau of the Census as well as other agencies.

These papers relate to one another so well that it is easy to forget that they are based upon the work of two different countries. Indeed, the papers furnish a good example of the principle of interchangeable parts. If Waksberg's paper had been presented by his Canadian counterpart, and Murphy's by his U. S. counterpart, the story would have been much the same, illustrating the effect of the long-continued close working relationships and interchange between Canada and the United States. This would not be equally true for the paper presented by Lee. In this case, differences in the scale of operations and in the resources available to users favor the earlier development of such a cooperative effort south of our common border.

A further reason for starting with Mr. Waksberg's paper is to be found in my own professional interest in evaluation of the quality of statistical data, as expressed in meetings of the International Statistical Institute beginning in 1951. The subject is one, moreover, with theoretical ramifications of considerable interest to many statisticians.

Dr. Taeuber's reference to President Washington's concern about the quality of the 1790 enumeration may have erroneously suggested a continuing attention to this problem. The fact is that for most of the long history of census taking in the United States, there has been an official tendency to take for granted that "if it's in the Census, it's right." One break in the long period of neglect - unfortunately, not benign - was that provided by General Francis A. Walker, Superintendent of the Censuses of 1870 and 1880. About one hundred years ago, General Walker, in words that might well have been written today, referred to the duty of producers of statistics to be candid about their shortcomings, even though as a consequence, the results of a census might receive less credit. However, his point of view was an exception to the prevailing one, and it was not until the midforties that the first post-enumeration survey was undertaken by the Bureau of the Census. Since that time it has been the standard practice of the Census Bureau, and of many other statistical agencies, including Statistics Canada, to regard the provision of measures of error in a census or survey as part of the task of production.

I believe that Mr. Waksberg's agency deserve much credit for introducing and maintaining this tradition of conceding the existence of errors, and seeking measures of their size and nature. Nevertheless, I am disappointed that we were not given more information on the quality of the 1970 Census. While it is true that the results of evaluation of the censuses prior to 1970 were delayed until well after the completion of other work, I had hoped that by the time of this meeting, about two and one third years after the date of the census, there might be provided a substantial amount of information about quality in 1970. Let us recognize that a great deal of work has been going on at the Census and hope that the completion of the evaluation task will be regarded as a high-priority undertaking.

Mr. Waksberg's report on coverage, based on earlier work by Jacob Siegel, was that underenumeration in 1970 had declined fractionally from the level for the preceding census. The improvement would have been about twice as great if the age-sex-color composition had remained unchanged over the decade. I would note also that the enumeration in 1970 took place after a bitter conflict which could easily have led to higher rates of undercoverage in 1970, except for the changes introduced in that year.

From my observation of the series of evaluation programs beginning in 1945, I am impressed with the increasingly sophisticated character of these undertakings. They provide a great number of different measurements, intended, as Waksberg has pointed out, to determine the effectiveness of the new 1970 procedures, to obtain another round of readings on the components of mean square error of important census statistics, to solve some problems of evaluation left from the 1960 work, and to pursue more intensively the correlates

and causes of error in census statistics.

Over the years in which the evaluation of censuses has been undertaken, there have been marked shifts in the relative emphasis given to the various parts of the evaluation program. For example, much less attention is now given to the reinterview survey as a means of determining the amount of error in a census. With the recognition that this device is subject to shortcomings similar to those of the original interview, dependence is now placed upon the analysis of demographic characteristics as a means of estimating overall undercoverage. The reinterview survey is retained, to be sure, but mainly to measure coverage of housing units and response variance.

Record matches are being employed to a greater extent than ever before, one of the major uses being to give a measure of reporting bias. The matches include many different files, such as Medicare registrants, motor vehicle registrations, Internal Revenue Service returns, expenditures for utility services, and prices received for houses sold. Each of these is used to provide evidence on the bias of one or more items in the census.

I was particularly interested in the report on the results of comparisons between the Current Population Survey and the Census. In the current comparison, as was the case ten years ago, the two sets of data were in quite close agreement. The situation was quite different in 1940 and 1950. In both those years, the census unemployment figures were significantly lower, even though the totals for the employed were in quite close agreement. The lower census figures for the unemployed were explained on the basis of the omission of considerable numbers of unemployed by the Census enumerators, who failed to ask all of the questions needed to identify certain types of unemployed. The most generally accepted explanation of the better agreement in 1960 and 1970 is that most of the population were given a chance to classify themselves, so that the numbers classified as unemployed more nearly equalled those obtained by the trained enumerators in the Current Population Survey.

Dr. Murphy's excellent paper describes an area of census work where changes have been unusually great in the past 20 years. After the 1951 Census in Canada, it was probably true that the great majority of all uses of the census depended upon the printed volumes, and relatively few upon other types of output. Even after the 1961 Census, if one may judge from experience in the United States, the amount of information provided in other than printed form was still not great. By 1971, however, the situation has changed sharply, with so many users having access to computers.

Nevertheless, it is at first surprising to hear that 90 percent of the prepared tabulations for the 1971 Census of Canada, the equivalent of 3,800,000 pages, will be in the form of micro-film, micro-fiche, or summary tapes. Even if only 10 percent is in conventional published form, the resulting 380,000 pages will be double that made available in the United States. The requirements for bilingual presentation in Canada would appear to explain only a small part of the difference.

I should like to commend Dr. Murphy for his statements on the responsibility of a dissemination program to give a description of the collection and processing procedures as they affect the data. He notes particularly that the imputation procedures may have a significant impact upon the results for very small areas, such as enumeration areas. I have wondered, therefore, whether his office plans to provide counts of the numbers of imputed cases for enumeration areas. Along similar lines, he has referred to plans to have the Dictionary eventually include measures of root mean square error. I raise the question as to how soon this will be done, and as in the case of the United States, would hope for all possible speed in getting out these measures, since providing information on the size of error should be regarded as an integral part of the production process.

Dr. Murphy has referred to three forms of publication: first, summary data; second, individual records, presumably with area detail deleted; and third, individual records manipulated so as to avoid disclosure. However, relatively little has been done apparently in the release of information under the second and third categories. In the United States, a great amount of information has been released in the form of individual records, so that we shall be much interested in the safeguards for such work adopted by the Canadians, who because of the smaller size of the country have a correspondingly greater problem in protecting confidentiality of individual records.

Dr. Murphy has told us of the great volume of preplanned publication tables which are listed in a new document called the Tabulation Directory. However, there will be some needs for tabulations not included in the Directory, and for these special work is required. I was particularly interested in his brief description of what might be called a "do-it-yourself" tabulation program in which the user willing to spend a little time in learning the system can use a special language called TARELA for a program designated as STATPAK. The need for this approach may be greater in

Canada than in the United States, where there are many summary tape centers to bridge the gap between the central office and the isolated user.

Finally, in Dr. Lee's interesting paper, we have a description of one kind of organization that has developed in the United States to serve needs that could be met only with considerable difficulty and delay by the Census Bureau. In Canada, where the provincial statistical systems are considerably stronger than their counterparts in the States, the corresponding decentralization of service to the public may take place in the provinces.

The potential services to be rendered by an organization such as Dr. Lee has described are quite great. It has been estimated that business has utilized no more than ten percent of the useful products of a decennial census, and there is little reason to believe that the academic and research users have done much better.

I recognize that I am less sensitive than I used to be to criticism of the Census Bureau, but it seems to me that the attitude toward the Bureau is more tolerant than it used to be. Indeed, the relationship between the organization described by Dr. Lee and the Census Bureau is described in a manner quite sympathetic to the Bureau. Occasional references are made to census errors and delays, but always in a context recognizing the size of the task of getting out census tabulations and tapes.

Dr. Lee's reference to the role of "Big Science" in the field of research and training seems to be well justified. The tapes at the data center of the Oak Ridge National Laboratory include all of those from the 1970 Census, as well as tapes from the 1960 Census, the Social Security 1% sample, County Business Patterns, the Census of Manufactures and a series of files from individual researchers. The development of appropriate software to use with these resources and of cooperative relationships with a number of social and physical scientists makes this new Center unusually promising for demographic research making use of new tools now available to social scientists.

One can wholly applaud Dr. Lee's statement that cooperation among institutions and the establishment of non-profit and semi-public data centers are needed for the proper exploitation of census materials. But the costs for many undertakings will be relatively high. There will be false starts and it will be necessary to write off some large outlays that fail to meet the expectations of their sponsors. But the potential yields are very great also, and we can be glad that organizations like the Oak Ridge National Laboratory have both the resources and the resilient attitude necessary to realize the research potentialities.