COVERAGE IMPROVEMENT AND MEASUREMENT
DISCUSSION

Cynthia Z.F. Clark, National Agricultural Statistics Service, USDA
Washington, DC 20250-2000

This session contains an interesting combination of papers investigating aspects of coverage in different types of censuses - population, housing, economic, and agriculture - with representation from both Statistics Canada and the Census Bureau. The three Census Bureau studies reported on use the capture/recapture dual system estimator to estimate coverage errors. The properties of this estimator were explicitly described in the 1986 JASA paper by Kirk Wolter.

The Lewis and Wolfgang paper on *Weighted Segment Research for Coverage Evaluation of the Census of Agriculture* investigates an issue of real import for the 1992 Census of Agriculture. Since 1945 the Census Bureau has conducted a coverage evaluation of the census of agriculture program. The 1987 and 1974 evaluations and the 1992 evaluations used the area frame sample of the National Agricultural Statistics Service. During the interim between 1987 and 1992, NASS changed its estimation procedures to use a weighted segment estimator in all States from the 10 States where it was used for selected commodity estimates in 1987.

The Census Bureau had the choice of taking advantage of this methodology and using the procedures for NASS's June Area Survey (JAS) or modifying the JAS procedures to use the open segment estimator. The later choice had cost implications, and did not provide the potential for increase in the accuracy of the estimates. To evaluate the impact of this change on the census estimate of farms not on the mail list, (which differs from the impact on the NASS area sample estimates) the study reported on by Lewis and Wolfgang was initiated. The study determined that the cost of that increased precision is an increase in bias primarily due to the "weight" used - the segment acres over the whole farm acres - with the accompanying difficulty of collection of the whole farm acres. In order to mitigate the bias, the Census Bureau and NASS have initiated procedures for 1992 to ensure that the whole farm acreage estimates are more accurate for use in the weighted segment estimator. The study also provides a basis for interpreting results from the 1992 evaluation as contrasted with the 1987 evaluation. This will be particularly valuable with the changes in data collection, processing, and estimating procedures.

The Childers paper on *The 1990 Housing Unit Coverage Study* presents a solid analysis of the census coverage estimator methodology as it applies to housing units as well as a specification of differences between the 1990 and 1980 sample designs. Both the P-sample, non-matches, and the E-sample erroneous enumerations, were analyzed by variables thought to contribute to these housing errors: type of enumeration area, occupancy status of housing unit, regional metropolitan census center, census region, size of block, housing units in structure, and race/origin of person 1. Investigation of significant differences found between categories of these variables should suggest changes and improvements in census procedures. The planned analysis of the 1990 person and housing unit coverage looks promising as it will facilitate investigation of duplicate persons and duplicate housing units as they relate to each other and to correct enumeration of both persons and housing units.

In contrast to the Childers paper, the Griffin and Moriarty paper on *Causes of*

Census Errors analyzed enumeration errors identified in the P-sample by variables associated with the respondent and the mode of data collection. Since approximately 75 percent of the census population was enumerated by mail, it is particularly important to examine factors contributing to self-enumeration errors. The factors examined included characteristics of the respondent, form length, timing of the mail enumeration, size of household, type of structure (multi- or single unit), and tenure status of housing unit. The surprising conclusion to the study is that self response results in a much lower erroneous enumeration rate (3 percent versus 9 percent). This difference is not attributable, however, to untrained enumerators. The characteristics of the individuals and housing units that do not respond to self-enumeration are the real source of the problem. The study also determined that questionnaires returned in a timely manner were less prone to include erroneous enumerations.

Both in the Childers and the Griffin and Moriarty papers, we learned a lot about the source, structure, and characteristics of errors in matching and enumeration of households and in enumeration of people. Errors were analyzed by variables that might provide differential errors within categories to determine which breakouts were significant. The next steps are 1) to analyze **why** a category of a variable might have a differential impact, and 2) to suggest and test alternative changes to census and survey data collection procedures that could reduce the effect on that category. In some cases, this process can be facilitated by the study of existing data as was done in the Griffin paper by looking at errors for mail self-enumeration versus personal enumeration. Differences were, in fact, found, and initial attempts

were made to identify different procedures to test such as residence rules and rostering procedures.

The Lorenz and Laniel paper on *Measuring the Quality of the Data on the Business Register: Methodology and Results,* discusses an ongoing program to provide measures of quality of the business register used for conducting both semi-annual surveys and censuses in Canada. The paper identifies specific content errors and coverage errors for each of the 18 industry divisions. A sample design is put in place that will enable periodic estimates of each identified error to be made. Table 3 of the paper is particularly insightful in that it shows the errors superfluous to specific industry divisions ranging from less than 1 percent (accommodations, food, beverages) to over 20 percent (mining), and a range of 0.0 percent (communication and also in government services) to 34.4 percent (educational services) missing from the appropriate division.

The measurements provided in the paper should, and have, enabled Statistics Canada to assess the impact of various errors in the register and to identify where resources could best be expended to improve the overall quality of the register. Again producing estimates of the errors is only the first step. The results need to be analyzed to improve the survey process - in this case, the development of the business register.

I would like to compliment the federal agencies for engaging in the kind of studies presented in this session and the authors for their thorough analyses of the structure of the non-sampling errors in the survey estimates under review. These studies should permit identification of major sources of errors and lead to steps to reduce the impact of those errors in future censuses and surveys.