DISCUSSION

Richard D. Burgess, Statistics Canada

The four coverage papers presented have addressed themselves quite effectively to particular but fundamental problems with population coverage evaluation methodologies. On the one hand we have the very complex difficulties of tracing and the consequent problems of dealing with unsuccessful tracing. On the other is the problem of adjustment or correction for coverage error which inherently leads to scrutiny of the limitations of the methodologies being applied. Indeed it may be said that the final test of the adequacy of the methodologies and procedures of coverage evaluation is whether adjustment from a statistical perspective is feasible and supportable.

The results of the paper on 'Small Area Adjustment' methodologies point out the difficulties in developing an adjustment approach in the face of incompleteness in the sampling frame - namely previously unregistered or unlisted elements of the enumerated and unenumerated population - and of not having a predictive model or knowledge of the distribution of undercoverage for smaller geographic areas.

In this regard it is essential to have for the province or state level supportable estimates of undercoverage in order to develop sound models for lower level adjustments. In this sense it must be possible to model the province or state level estimates based upon national level estimates. In general no model of undercoverage, based upon characteristics of the population or of households or dwellings, developed to date has led to an explanation or a prediction of differences in undercoverage from one geographic area to another. As an example, synthetic estimates for provinces, which are based upon national estimates of undercoverage for various population characteristics provided by the Canadian Reverse Record Check, do not conform well to the Reverse Record Check estimates for provinces. I see the same general type of problem in the U.S.B.C. results. As is shown by these results, the problem is confounded by potential limitations of the available aggregate estimates.

For the studies discussed in the papers, representativeness is always a problem. This is particularly true for the CPS and IRS samples used in the experimental matching studies. They do, however, provide insight into tracing, matching and non-response methods and problems, which is the primary intent here.

The possibility of using the IRS file, in particular, to assess the completeness of an RRC type independent list should be investigated. The two groups cited - illegal aliens and immigrants - would have special importance in this regard. The IRS file is of further value in providing up-to-date name and address data for an undercoverage study providing some ready link to this file is possible.

Where tracing is required coverage evaluation study results have the potential to be significantly biased if failure to trace occurs for a significant proportion of cases. A particular set of approaches to tracing have been proposed in the paper on 'Forward Tracing'. I will agree that a ten year period between censuses improves the independence aspects on an RRC and that it also makes tracing an extremely difficult task. I must conclude that the best hope of tracing people, with forward or retrospective tracing lies with easily accessible lists of people - with their addresses - motor vehicle and IRS records are examples. Carried to the ideal such lists would constitute an up-to-date, even if fragmented and repetitious, list and location of the total population.

My primary concerns with the forward tracing approach are cost and respondent conditioning. If there is no contact with the sample person, then all persons in the sample must be traced periodically and this will be expensive. Admittedly the retrospective tracing, because of the ten year period, may be as expensive and much less timely than the forward trace approach. The cost will depend upon the exact nature of the tracing operations which, understandably, is not detailed in the paper. The cost of forward tracing without personal contact might be reduced and the tracing made more efficient by initially matching, using name and address - manually if necessary - to some of the administrative files cited in the paper. For example, a match to the IRS file would yield a Social Security number of the selected person or of a spouse or parent which could be used years later for tracing. The same effect, of course, could be achieved by the initial contact approach. Under this scheme some of the people will not provide the information or will not be located, thus there would remain some need to refer to an administrative file.

Forward tracing with annual contact is very likely to have some conditioning effect. This is the rationale behind Census publicity, although forward tracing will not necessarily have a positive bias. The possibility of conditioning with only the initial contact is clearly less of a problem. Even if this is so it is not clear that an assessment of conditioning can be made in the proposed experiment - either between the four approaches or with past experience. The conditioning effect will not be examined over the full ten year period. (This will also impact on the assessment of cost.) Further, it is not clear how household (occupied dwelling) undercoverage will occur in the experiment and what the impact of this will have on analysis or conclusions. The conditioning potential may have to be examined in an actual census.

No trace rates for the '76 Reverse Record Check are cited in the 'Forward Trace' paper. The results for the '81 Reverse Record Check are not altogether similar to those for '76. The overall no trace rate decreased from 4.8% ('76) to 3.4% ('81). The change for the Immigrant frame (10.6% in '76 to 3.9% in '81) and for the Missed frame (9.6% in '76 and 3.9% in '81) were dramatic. For provinces the overall no trace

rate was as high as 7.7% in 1976 but no higher than 5.1% in '81. While it may be concluded that the '81 Reverse Record Check was more successful in tracing its sample of persons it may be more important to address the potential for unpredicability in these rates and the certainty that there will always be no trace cases.

It is clear from the examples in the papers presented that the no trace or non-response cases have a higher probability of being truly missed than what would be suggested by using all completed cases indiscriminantly. In the '81 Reverse Record Check less than two-thirds of the sample was used in imputing or weighting for no traces. Further, this process was carried out within age-sex and geographic groups. The result of this was to impute a missed rate for the no traces 60% higher than that for completed or traced cases.

Did we go far enough along the lines suggested in the paper on the 'Matching Studies' and discussed in some detail in the paper on 'Missing Data'? The latter paper makes it clear that using all completed cases is an inappropriate approach. It also makes it clear that this is not the real problem at hand. What is required is to identify specific cases to be excluded or included in the imputation process. A complex tracing and follow-up system with repeated operations should yield a logical structure which can be used to make this determination. In the end, however, there will remain an information gap which will not permit elimination of the question of bias. However, with extension of the approaches discussed to a detailed level it should be possible to categorize data to a sufficiently refined level that assumptions have little impact on final undercoverage rates.

At the same time in so doing it is necessary to guard against attributing to no trace cases characteristics which are merely a function of tracing steps and procedures and on the other of these steps. There is perhaps some increase in this risk as the time between the list creation and census day increases.

In general I have found the papers interesting, thought provoking and to be of considerable merit. The work is well worthwhile and I believe developments on all issues in this area must be continued.