DISCUSSION

Richard Platek, Statistics Canada

Rather than comment paper by paper, I will touch upon some of the more interesting points covered collectively or individually by the papers. I would like to voice my appreciation to the chairperson and to the authors for having organized and contributed to a very cohesive and interesting session. I am particularly impressed with the experimental work being carried out by the U.S. Bureau of the Census into effects of different data collection methodologies on survey estimates and operations.

The Current Population Survey (CPS) and the Canadian Labour Force Survey (LFS) are in general very much alike. Both produce seasonally adjusted monthly data on current labour market conditions, collected through personal or telephone interviews. Both use stratified multi-stage designs with ratio estimation. Also, both surveys are generally redesigned every ten years after decennial censuses.

Both organizations, the Bureau of the Census and Statistics Canada are at such a stage now. That is, both are after their respective decennial Censuses and projects and plans are well under way aiming at redesigning the CPS and the LFS, along with other major household surveys. Various aspects of these projects and plans are well documented in the two papers on redesign activities presented here today. Both organizations can benefit from each others experience. We, in Canada, evaluate the results of your studies very carefully. Let me start by highlighting some of the common features yet different approaches used in the two surveys together with some problems associated with them.

Sampling Frames. It may be noted that both surveys use a combination of area and list frames. However, they differ greatly as it now stands in the degree and level at which the list frame is used. In the case of the LFS, the lists of dwellings are prepared and maintained for sampling purposes within final stage areal sampling units (clusters), which are generally the size of blocks or smaller. There is no dependence on the census or any other external sources in this approach. On the other hand, the CPS approach rests heavily on the Census and other external sources in creating as well as updating its list frame In fact the CPS uses a multi-frame approach.

There are a number of reasons for concern about multiple sample frames using lists from various sources over which there is less than satisfactory control. One problem is controlling overlaps. Some types of households may get double chance of selection which may obscure the fact that other households have virtually no chance of selection i.e., are missed somehow. This is difficult to measure and control. There is also the problem of dependency on sources which are not directly controlled. This is particularly true of building permit information. It may sometimes be more expensive to control and maintain such frames over the life of a survey design than to set up a special frame even though the initial per unit cost of the special frame may be higher. In the case of sampling frames, such requirements as completeness, stability, ability to control overlaps, efficiency in updating, unambiguity, all must be taken into account.

General Purpose Survey

Another interesting problem is the case when several surveys share the same frame and/or data collection facilities. Should one optimize the design for a dominant survey, then do the best one can for the rest, or, should one try to optimize over several surveys using the same facilities. It has to be pointed out that it is difficult enough to ascertain the objectives and priorities when one survey is involved. It is often impossible to do this simultaneously for a number of surveys. Of course, another option is to design a general purpose frame which will serve as a base for the design of the individual surveys. Cost benefit analysis and uses of such a survey vehicle is being studied by both organizations.

Alternate Rotation Pattern: Research activities are being undertaken for both surveys in relation to the appropriateness of different rotation patterns.

In a rotating panel survey such as the LFS and CPS, the monthly sample size determines the reliability of monthly estimates of levels and rates. However, it is primarily the rotation pattern which determines.

(i) The reliability of estimates of change, whether month to month, quarter to quarter, or for a calendar month from one year to the next and

(ii) The reliability of estimates obtained by combining monthly data for quarterly, semi-annual or annual estimates.

In general, rotation patterns which result in more reliable estimates of change are less reliable for estimates of levels and vice versa. The correlation between samples in successive months are advantageous for estimates of change but are disadvantageous for estimates of levels. For example, the current LFS rotation pattern under which households remain in the sample for 6 consecutive months and then are permanently retired, when compared with the rotation scheme used in the CPS is not as efficient for quarterly or annual average estimates of level.

On the other hand the rotation scheme used by the CPS is less efficient than that of the LFS for estimates of change. Thus, the choice of a rotation pattern should be governed by the relative priorities attached to these types of estimates.

Stratification and Sampling

A small but important point regarding stratification especially due to infrequency of major

redesigns which take place only after 10 or in some cases 20 or more years, is the stability of stratification variables over time. It is usually wise not to stratify too deeply and to give considerable prominance to geographical stratification. As regards the choice between the two alternatives namely, one or two PSUs per stratum it would be of interest to compare these alternatives for various sizes of PSUs and strata and for different stratification criteria. For instance, if the gain due to stratification were insignificant would 1 PSU per stratum be better than 2, and if so to what extent, and how the choice would impact on the problem of estimating variances.

In the Statt et al paper, in the choice of an optimal month to begin sampling permits, the model used to estimate the optimal month (k) such that the numbers of permit lag units and duplicate units are at a minimum seems to be quite an appropriate one, and appropriately several types of biases that may arise in the process have been identified. However, there is no discussion on the compensation for such biases as a result of omissions which may arise even with the optimum month. In addition, the optimum value of K when rounded up to the next higher integer, as suggested in the paper, will give rise to certain degree of bias. It is not unlikely that this bias may become quite significant in certain situations. In fact the lower the variance of K the higher would be the effect of rounding up. Another question that arises is how flat is the optimum? But this can only be answered by examining the distribution of individual cell levels.

Another point to be noted is the variation in the construction activities from area to area and from one period to another. In order to account for this it may be desirable to consider more areas, subject to operational constraints and more importantly to have the value of K updated from time to time. Data based on 1974 may already be out of date in certain cell levels.

Recall Bias Study in the National Crime Survey

John Bushery's paper has very clearly identified the scope of the experiment and its limitation. The conclusions arrived from the experiment are what one would normally expect, namely that the level of crime reporting decreases as the length of reference period increases. Reference period of 3 and 12 months have been tested against the current 6 month.

From this, one may infer that the reference period of even less than 3 months could result in improvements due to further reduction in recall bias.

Since the paper describes the result of the experiment, the high emphasis given to bias and mean square error does not seem to be out of place. However, in using the results of this study in practice one has to examine the entire strategy and the overall survey operations. Factors to be considered at that stage would include the relative cost of survey, i.e. for example, would it be desirable to reduce the sample to about 5000 households from currently 10,000 households per year. No doubt current survey provides very low CV at the national level but what would be the effect of reducing the size of the sample particularly at the state or Regional levels. It may be noted the bias (B_3 or B_6) would not normally be affected by the sample size.

If 3 month reference period is adopted instead of the current 6 month reference period, the different panels will be interviewed 12 times during the 3 years in the sample. This itself may have adverse effect on the response rate and induce response errors due to respondent fatigue. One possibility would be to keep the panels in the sample for a period of 2 years so that only 8 interviews would be required. This may however, result in an increased cost and loss of longitudinal data for one year.

Thus choice of reference period need to be examined in the general context of entire survey operation objectives and cost.

Finally, the Abramson et al paper is well and confidently written urging other survey organizations to follow the same methodology, i.e., to select address from permit address listings for interview purposes. My only comment would be that I would like to see more detailed examination of alternatives. I feel that the alternatives examined are really a variation of the same theme.