DISCUSSION

Joseph Waksberg, Westat, Inc.
1650 Research Blvd., Rockville, MD 20850

1. Drew-Dick-Switzer Paper

The set of research projects on the efficient utilization of telephone survey methods for the Canadian Labor Force Survey described in this paper and in earlier ones by some of the same authors and other collaborators in Statistics Canada is very impressive. The priorities in choosing the aspects of methodology for examination were obviously carefully considered and, I think, basically correct, as are the study designs and analysis of results. I look forward to seeing how the results of the research are implemented in the revision of the Canadian Labor Force Survey.

In listening to this report and comparing the results to those in Clyde Tucker's paper as well as my own and other U.S. experience, I was struck by how difficult it is to establish general truths in this field. Problems appear to be unique and have to be considered in the context of the population covered and the subjects studied. There are, in addition, important differences between the U.S. and Canadian situation. Let me give some examples.

a) An important part of the research relates to the methods of integrating telephone samples with area samples. Such integration is needed because some households do not have telephones. Statistics Canada believes that nontelephone households need to be included even though the telephone coverage is over 98 percent. In the U.S., 98 percent is higher than the coverage rates achieved in such surveys as CPS and the NHIS which are thought of as relatively unbiased samples of the total population. If I were at the Census Bureau, I think I would consider myself lucky to get 98 percent coverage through a cost-efficient system such as telephone sampling. Apparently, it is possible to do better in Canada and this shapes the methodological requirements.

b) The response rates obtained in cold telephone calls in Canada appear to be much higher than those obtained in U.S. surveys. This is true for those attempted by the Census Bureau and other U.S. Government agencies as well as surveys carried out by non-governmental organizations. One obvious possible explanation is that the Canadian potential respondents are more cooperative than those in the U.S. However, it would be useful to examine the approaches used by Statistics Canada to induce cooperation. Perhaps they use techniques that have not occurred to us.

c) The omission of nontelephone households from the sampling frame will introduce serious biases in the statistic for some items and have a negligible effect for other items. The unemployment rate in nontelephone households in Canada was triple the rate in telephone households. This wide diversity in rates is similar to what Westat found for school drop-outs (Burke, 1988). The drop-out prevalence rates for persons 14-21 years of age were 31 percent for those in nontelephone households compared to 7 percent in telephone households. A purely telephone frame for surveys on these topics is clearly out of the question. On the other hand, Westat found only modest biases in estimates of other items relating to education, for example, the proportion of preschool age children involved in school programs. Massey and Botman (1988) show data on a variety of health characteristics in telephone and nontelephone households. For many of the items, telephone and nontelephone households are quite similar, but there are striking differences for items that are income-related, such as availability of private health insurance or frequency of dentist visits.

I should note there are also similarities between the U.S. and Canadian experience. I will give two examples.

a) The authors report striking differences in unemployment rates between members of households which have newly assigned telephone numbers and other households. In Quebec, the unemployment rates were 20 vs. 11 percent, and in Ontario 10 vs. 7 percent. The exactly equivalent rates in the U.S. are not available, but a reasonable approximation is the rates for mobile persons and those who have not changed their residence. The most recent CPS report on mobility (Census Bureau P-20, No. 425) shows an unemployment rate of 11 percent for movers and 7 percent for nonmovers, almost identical with the Ontario figures.

b) The experiment with CATI showed a small improvement in within household coverage with CATI operations. Maklan and Waksberg (1988) report that in a group of RDD telephone surveys conducted by Westat there was evidence that within household coverage was slightly better than CPS. The improvement was not attributed to CATI in the 1988 report, but the Canadian experience makes that a plausible explanation.

I'd like to add a few other comments on the paper.

In regard to the substantial difference in unemployment rates between telephone and nontelephone household, it was indicated that omission of nontelephone households would reduce the estimate of the unemployment rate by 0.2 percent. It would be worth examining whether the noncoverage bias could be reduced significantly by appropriate weighting, somewhat similar to the use of weighting for nonresponse adjustment. It's worth noting, however, that Massey and Botman's (1985) report on noncoverage for health items showed only a slight improvement with weighting. This was also true in Westat's examination of drop-out statistics.

The Canadian Labor Force Survey is a panel survey, as are many of the well known Census Bureau household surveys. It's not clear to me how panel operations would be handled in a telephone frame. The most direct system is to consider the sample as one of persons, and try to locate and interview movers during the life of the panel. This would improve the analytic capability of the panel features, but would probably result in an increased nonresponse rate. Alternatively, it could be considered a sample of addresses with personal visits made to contact new residents of the sample addresses, as is done now. Reference was made to lists of newly activated telephone numbers. Is the intent to use a sample of these numbers as replacements for inactive numbers, and if so does this provide a reasonably unbiased sample.

The authors indicate that centralized calling for the 2nd and later months in each panel is a likely direction for the Canadian LFS. This, of course, will reduce the field interviewers' workload and may require reconsideration of the field organization. It is usually considered important to provide a decent workload to each field interviewer. It improves the efficiency and helps in retention of interviewers. If the telephone interviewing is taken out of their hands, an optimum sample design may call for a smaller number of PSU's. The effect on the variances may not be increased very much if the PSU's are made larger in size. Since most of the interviewing is done by telephone, the increased travel in larger PSU's should not add much to the total cost.

2. Clyde Tucker's Paper

The impetus behind Clyde Tucker's examination of lists is the desire to improve response rates in telephone surveys, by mailing advance letters to potential respondents. Since RDD does not provide mailing addresses, commercial telephone lists are used for part of the sample. I think the emphasis on response rates is correct. Almost all RDD surveys I have

looked at seem to have poorer response rates than equivalent face-to-face interview surveys. Research which sheds additional light on the causes of nonresponse or suggests methods of reducing it could thus have an important impact on survey methods.

Clyde provides interesting and useful information on the characteristics of the list frames, and on the problems of working with them. These problems include the completeness of the lists (or their incompleteness), duplication, occasional lack of mailing address, and other identification problems that interfere with geocoding. Another problem in working with lists as a frame that Clyde does not discuss is that households with two or more telephone numbers will need to be queried about all telephone numbers used in the household, and the additional numbers searched for on the lists.

Given all of these problems, I wonder whether it's worth spending any more time considering a dual-frame approach. The main reason for considering the lists is to take advantage of mailing addresses by sending advance letters. However, this can be done in the context of a single frame RDD survey, with the potential telephone numbers in the sample clusters matched against the lists and the matched cases sent advanced letters. This avoids any need to eliminate duplication, concerns about the completeness of the lists or possible biases due to errors in matching the RDD frame with the list. It also increases the sampling efficiency of the design because with a dual frame design the probability of selection for households on the list will be different from those not on the list. Because only about 55 percent of telephone households are on the lists, the increase in variance with a dual frame design over a single frame sample is fairly high. For example, if equal sample sizes were selected in the RDD and list samples, the increase in variance would be about 30 percent. If the list sample was twice as large as the RDD sample, the increase would be about 70 percent. Considering that the success rate in getting a household with the list sample is 85 to 90 percent, and the rate in RDD with the Mitofsky-Waksberg procedure is 60 to 65 percent, it is highly unlikely that the reduction in cost of screening is large enough to make up for the increased variance. This is even without taking into account the potential biases due to errors in matching, duplication, etc.

I also have a few comments on some of the details of Clyde's paper. First, it's useful to recognize that some of the issued raised relate to the use of a telephone survey to identify establishments that will later be visited in person. The later need for personal visits forces the telephone survey to be restricted to a sample of PSU's. This requires grouping the telephone numbers in the frame into PSU's. The PSU structure applies to both an RDD and a list frame. Clyde pointed out the difficulty of accurately associating telephone numbers with counties, and that a geographic screening question was ultimately necessary to exclude respondents outside PSU boundaries. It may not be necessary to put so much effort into getting an exact geocoding scheme. One could have an unbiased procedure of associating telephone prefix numbers with counties, and defining a PSU as the set of prefix numbers corresponding to the counties in the PSU. There would be some loss of efficiency because the measures of size would not correlate as highly with the PSU populations as in a normal area sample. However, this loss could be compensated for by the reduction in geocoding activities and the possibility of mistakes in the geocoding operation. Using a set of prefix areas as the PSU should not affect the interviewer travel costs appreciably.

I have one other comment regarding commercial lists of telephone numbers, which contains names and addresses. They can be used for survey operations in other ways than as a means of sending advance letters. Other possibilities are:

a) The addresses can be geocoded to Census tracts or other small areas to permit estimates of socio-economic characteristics to blocks of telephone numbers. Telephone numbers can then be stratified prior to sample selection. Mohadjer (1988) describes a use of such stratification to oversample minority groups.

b) Similar matchings of addresses with Census geography can assist in geocoding, when the sample is constrained to be in PSU's, or is restricted geographically in other ways.

c) The telephone lists can be used to establish measures of size for banks of telephone numbers, so that selection with probability proportionate to size can be carried out. Cassady's paper describes this sample design in his classification of hybrid designs.

These comments shouldn't distract attention from the main purpose of the research, which is to test the ability of advanced letters to improve response rates. Studies on the effectiveness of advanced letters were reported at the 1986 ASA meeting by Groves and Lepowski and by Drew and Jaworski with somewhat conflicting and puzzling results. I hope the new study will shed further light on the possibility of improving response rates, and look forward to seeing the results.

3. Robert Cassady's Paper

Bob Cassady's paper seems to me an elegant way of organizing the mathematics of the various sample designs available for telephone surveys so that essential features that are common, or different, among the designs can be examined. Bob Cassady has used it to calculate costs and variances and a measure of the efficiency of the various designs.

I went over his expressions for costs and variances prior to this session more carefully than can be done while listening to a paper at a meeting, and the derivations appear to be straightforward and correct. However, there is one assumption in his derivation of the variance of hybrid design that understates its variance. He assumes that the measures of size, which are based on the number directory listings per cluster, are perfectly correlated with the number of households, so that a self weighting sample provides a constant number of households per cluster. In practice the correlations will be far less than one, and the variability in cluster size will add a component to the variance. The variance of the hybrid estimate will thus be larger than is shown in his paper. This also applies to the stratified estimate. Drew and his associates mentioned the fairly rapid deterioration of the quality of the directory frame. It is likely that an equivalent deterioration occurs in the U.S., and this could seriously affect the correlations unless the frame is almost continuously updated.

Sampling theory doesn't provide much insight into the effect of variability of cluster size on the variance when poststratification is used. (There can be a sizeable increase for unbiased estimates, that is when the weights reflect only the reciprocals of the probabilities of selection, but telephone surveys usually use some form of poststratification.) The effect of variability in size depends on whether there is any correlation between the sample size in a cluster and the variable being estimated. I suspect that the correlations are fairly low for most variables commonly used in surveys, but there could be big differences among surveys, or among variables in a single survey. It would be useful if researchers using the hybrid sample design prepared estimates of the variability in cluster size, as well as the possible impact on the variances.

Let me make a few comments on the efficiency of the designs that Bob examined. Bob cautioned that one should not generalize too much from his one set of assumptions on costs and intraclass correlations. I would like to underscore the caution. The comparison of costs is very sensitive to the assumption one makes of the ratio of c_p/c_u, that is the cost of a productive to an unproductive call. He used a value of 2 for his

illustrative example. The ratios are much higher in most of our work; ratios of 5 are more common, although the ratios are very dependent on the size of the questionnaire and other survey conditions. With a ratio of 5, the costs of the Mitofsky-Waksberg and the hybrid designs become much closer, as do the variances because with a ratio of 5 the optimum cluster size is about 2 or 3 rather than the 5 used in the paper. In fact, the efficiency of all designs except the single stage sample become fairly close. With such a situation, it is useful to consider features of the sample designs other than cost and variances in choosing among the alternatives.

In going over the five sample designs discussed, I would be inclined to dismiss three. (a) The single stage design is clearly much more costly than any of the others, under reasonable assumptions of the value of c_p/c_u. (b) I am dubious of the use of the stratified design. With a value of c_p/c_u of about 5, it is not importantly cheaper than the Mitofsky-Waksberg design, and much more complicated to implement. Further, I agree with Bob that the proportion of telephone households in the hybrid frame will be very high, probably .97 or .98 percent or even higher. If one is willing to accept the 93 percent of telephone households as a reasonable approximation of the total population, it's hard to see why one should hesitate about accepting another one or two percent loss in the coverage. (c) The dual frame sample is complicated, with no important cost savings. It also introduces many of the problems described by Clyde, that is duplication, boundary problems, and the need to match the RDD and directory samples. There are also potential problems in weight adjustments for households with multiple telephones.

This leaves the Mitofsky-Waksberg and the hybrid design as the principal choices. There is one very attractive feature of the hybrid design; it avoids the sequential sampling called for in the Mitofsky-Waksberg plan. The sequential aspect requires careful attention and control of the operations. More important, it is time consuming, making it difficult to implement when there is a tight time schedule. I hope Bob Cassady and his colleagues continue their research on the hybrid design, particularly on the cost of purchasing the required data, and the variation in cluster size and its consequences. The research should also examine the effect of the aging of the measures of size. As a practical matter, most research organizations would not expect to continually update the measures of size but purchase a new list every 2 or 3 years. (I assume the cost is non-trivial.) Would this introduce an appreciable bias or significantly change the variability in measures of size.

There is another class of designs coming out of Potthoff's generalization of the Mitofsky-Waksberg procedure (Potthoff 1987) Cassady does not analyze the costs or variances of sample design variants in the Potthoff system. One of the purposes of Potthoff's generalization is to reduce the amount of effort devoted to the sequential aspects of sampling. I would be interested in comparing the effectiveness of several designs coming out of the Potthoff sampling procedure with the hybrid design.

I have one final remark on these sample designs. A useful feature of the Mitofsky-Waksberg design is that it achieves the exact sample size one specifies. With the hybrid and Potthoff designs, one can specify the expected sample size, but the sample sizes for most surveys are random variables and will differ somewhat from the expected value. From the point of view of precision of the estimates, there will be very little difference. However, a company like mine responds to RFP's from U.S. Government agencies, and many of them specify the exact sample sizes required. We're not always sure we can explain to the contract officers of these agencies the difference between actual and expected sample sizes and why the difference is unimportant. It is comforting to have a design which will provide the exact sample size required.

REFERENCES

Burke, J. (1989), Undercoverage Bias in the Field Test for the National Household Education Survey, unpublished report to National Center for Educational Statistics by Westat, Inc.

Drew, J. P. and Jaworski, R. G., Telephone Survey Development on the Canadian Labor Force Survey (1986), ASA Proceedings, Section on the Survey Research Methods

Groves, R. M. and Lepkowski, J. M., An Experimental Implementation of a Dual Frame Telephone Sample Design (1986), ASA Proceedings, Section on Survey Research Methods

Maklan, D. and Waksberg, J. (1988), Within Household coverage in RDD Surveys, Telephone Survey Methodology, John Wiley & Sons, New York

Massey, J. T. and Botman, S. L.(1988), Weighting Adjustments for Random Digit Dialed Surveys, Telephone Survey Methodology, Wiley & Sons, New York

Mohadjer, L. (1988), Stratification of Prefix Areas for Sampling Rare Populations, Telephone Survey Methodology, Wiley & Sons, New York

Potthoff, R.F. (1987) Generalizations of the Mitofsky-Waksberg Technique for Random Digit Dialing, Journal of the American Statistical Association, Vol. 82, No. 398, June 1987

U.S. Bureau of the Census (1988), Geographic Mobility, March 1985 to March 1986, Series P-20 No. 425