

Joseph Waksberg, Westat, Inc.

1. Subjects Discussed in Papers

The papers presented mostly focus on two subjects -- attempts to improve the response rates in telephone surveys and to increase the effectiveness of the sampling operations. I was interested in seeing that the papers do not indicate any real concern about the quality of responses in telephone reporting; the issue is just briefly mentioned. This lack of concern reveals something about our current state of knowledge, or perhaps, beliefs. It is somewhat reminiscent of the famous Sherlock Holmes story in which the clue was the fact that a dog didn't bark. I think that most survey practitioners, myself included, believe that there is ample evidence that the quality of reporting is not much different in telephone surveys from those in which interviewing is conducted on a face-to-face basis. There are, of course, differences in the ways that some questions need to be worded and questionnaires formatted in the data collection methods. However, the major research needs that are unique to telephone surveys relate to other aspects of survey design.

I think emphasis in the research on response rate control and more effective operations is appropriate. Almost all RDD surveys I have looked at seem to have poorer response rates than equivalent face-to-face interviews studies. The lower levels have inhibited the use of RDD in several major, continuing U.S. Government surveys. Any 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.

Development of new methods of improving the sampling operations is also an area that needs examination. The Mitofsky-Waksberg sampling procedure which is now commonly used for RDD studies has largely eliminated the need for the excessive screening of nonworking numbers that existed in older methods of selecting unbiased RDD samples. However, the Mitofsky-Waksberg method is a partially sequential procedure with stop rules, and this unfortunately creates the need for a fairly lengthy interview period, if a decent response rate is desired. There have been a number of attempts to reduce the elapsed time required (other than by simply using list samples). The methods suggested to date have some undesirable features, and I am glad to see imaginative new options being developed.

Since almost all of the topics discussed are touched in a number of papers, I think it is more useful to organize my comments by subjects, rather than discussing each paper separately.

2. Response Rate Improvement Through Advance Letter

The two papers by Groves and Lepkowski and the one by Drew and Jaworski examine possible improvements in response by sending advance letters to the sample households. Names and addresses are, of course, not available with a pure RDD sample. As a result, a combined RDD-list sample needs to be used with lists of telephone subscribers purchased from a commercial organization. The listings are also used for sample design purposes but, for the moment, I will concentrate on their use in reducing nonresponse.

Both the Groves-Lepkowski and the Drew-Jaworski studies report useful improvements in response rates, although there are some puzzling differences in the details. The differences may, of course, be due to the fact that the

populations studied in the U.S. and Canada are not the same, but it is useful to look into them to see if they reveal anything else.

I should note that there seems to be an implicit assumption that a list of telephone subscribers is necessary for advance letters to be mailed. Westat tried a somewhat different approach a few years ago with an RDD sample. In the first contact with a sample household, we only told the potential respondent that we wanted to send him or her information about the survey, and asked for name and address. The advance letter was then sent out, followed about a week later by a telephone call to elicit the required information. The procedure had no perceptible effect on response rates. Apparently, whatever advantages there were from the advance letters were negated by refusals to provide names and addresses. Using a combined list sample-RDD approach gets around this problem.

The increase in response rate reported by Groves and Lepkowski is about the same order of magnitude as reported in the Drew-Jaworski experiment. However, if we look at the breakdown of the nonresponse, it can be seen that refusals only account for half of this difference, with the rest coming from not-at-homes, language problem, etc. Should the advance letter affect reasons for nonresponses other than refusals? The increase in the other nonresponses may reflect differences in the List population and total telephone population. This would be consistent with the Groves-Lepkowski findings.

It would be interesting to see identical breakdowns for the two studies, that is to see separate data for refusals and other nonresponse in the Michigan study and a breakdown of the RDD sample into the part that appears on the purchased list and the nonlist part. Would that be possible?

There is one other aspect of the two studies that should be noted, the very large differences in response rates, even when approximately the same procedures are used. For RDD, the Michigan study shows about a 60 percent response while the Canadian one is 90 percent. I assume the difference between populations covered by the two studies and the name-recognition and prestige of a Government agency in Canada are largely responsible for the high level in Canada. I am not sure that devices for improving response that are effective when response rates are in the neighborhood of 60 to 70 percent are useful or necessary when responses are close to 90 percent. However, the Michigan experience seems unusually low. If 65 to 70 percent is the best that can be done, even with an advance letter, much more research is necessary to investigate other methods of improving response rates.

Effect to Short Interview Period on Response Rates

The Census Bureau report has interesting information on the effect of a short interview period on response rate. Extending the interview period from the original two weeks by about another two weeks, produced useful improvements in response rates. Flemion and his colleagues state that the extension increased the rate by three percent. I am curious why the response rate of the extended period in January-March 1986 was compared to the rate in April-December 1985 rather than to the one in January-March 1986. The latter shows a six percent increase. This seems to be a more appropriate comparison. It is also more consistent with our general experience at Westat. The fact that Sunday is a

productive day for interviewing, with only a trivial effect on refusals is also similar to what we find.

I am puzzled over the reports at both the Census Bureau and Canada about the much higher nonrefusal nonresponse rates for RDD than in face-to-face interviews. These rates are essentially due to persons who are at home only infrequently, those who are away for extended periods, have language problems, or are ill. The latter causes should affect both RDD and face-to-face interviewing in the same way. For those who are hard to get hold of, RDD permits more call backs and probably calls later in the evening. Are the two rates really being calculated in the same way, or is there a practice of including uncertain telephone numbers, which are mostly nonresidential, with nonresponse. In calculating response rates, we generally estimate a proportion of the unresolved sample cases that are likely to be residential, rather than assuming they all are.

For studies in which we think it important to have quite accurate estimates of response rates, we try to clean up unresolved cases by checking with the business offices of the local telephone companies. We have usually been able to obtain the exact status of about two-thirds of the cases that could not be disposed of by other means. This has not improved our response rates but has given us a better measure of it. Our experience is fairly similar to the Census findings -- a little under half the cases are residential.

3. Sample Design Issues

Listings as Measures of Size

Both the Michigan and Canadian studies use cluster samples, with clusters defined the same as in the commonly used sample design for RDD. However, unlike the Mitofsky-Waksberg method, both studies rely on outside data for measure of size. The Canadian study relies completely on counts supplied by others, and the Lepkowski-Groves experiment relies partially on counts of listed households. There are important advantages to using predetermined measures of size, and some possible problems.

Both papers point out that the amount of screening is reduced. The need for the first-stage screening is completely eliminated in the Canadian plan, and partially so in the Lepkowski-Groves method. The ability to use listed numbers for advance letters is also cited. Another advantage, which is not mentioned in either paper but which may be at least as important as the one cited, is the fact that using outside data for measures of size avoids the partial sequential operation required for a self-weighting sample in the Mitofsky-Waksberg procedure. The sequential aspect has some convenient properties -- it guarantees exact sample sizes -- but it creates the need for long interview periods and detailed record keeping. With predetermined measure of size, the within-cluster sample is changed from one that calls for specific sample sizes to one with fixed within-cluster sampling rates. The latter is much simpler to apply and more importantly can reduce the data-collection period.

As far as I know, in the U.S. we do not have the ability to obtain data on the total number of households per cluster, as appears to be the case in Canada. The measures of size used by Lepkowski and Groves thus reflect listed units only. There is a possible reduction in efficiency from this feature of the sample. The sample does not have equal cluster sizes as is the case with the Mitofsky-

Waksberg design. Since listed number comprise only about 60 percent of the total frame, I assume the correlation between measure of size and the actual number of households is only moderate, resulting in variation in cluster size. I think it would be useful to explore the extent of this variation and its possible impact on variances.

In this connection, there is one statement in the Drew-Jaworski paper that puzzled me. The description of the RDD samples says that a sequential procedure is used to assure getting 1 to 2 residential numbers per PSU. With pre-assigned measures of size, a sequential plan should not be necessary. In fact establishing fixed sample sizes per cluster will introduce the need for weighting, or could have some biases. Did I misunderstand part of the design?

The Groves-Lepkowski and Drew-Jaworski papers used geographic stratification in sample selection which, of course, can be done fairly readily in RDD surveys. Westat is starting to examine the efficiency of additional stratification. Donnelley has prepared a tape for us that provides a variety of 1980 Census characteristics of prefix areas. They have done this by examining their file of addresses and telephone numbers, associating prefix area with zip codes and using the zip code tabulations of the 1980 Census. We have carried out some small-scale tests of the accuracy of the data with reasonably satisfactory results. We plan to use the tape for a large scale survey starting in October, and should have information on its effectiveness in 6 to 12 months. The purchase cost is fairly high and it's uncertain whether it is cost-effective when it is used solely for stratification. Our current plans are to use the tape as a vehicle for oversampling rare populations such as Black, Hispanic, low-income, etc. We may be in a position to report on its usefulness next year.

Dual Frame Samples

Dual Frame samples are involved in the four projects. Groves and Lepkowski combine list and RDD frames. The Drew-Jaworski paper discusses both list and RDD frames and telephone and area sample frames. The Census paper does not contain any examination of telephone and area samples but I assume it will be necessary to combine the two if RDD is made part of the data collection methods.

The Michigan study showed a quite large difference between the number of telephone households who report themselves as being listed and those who are found to be listed in a matching operation. Presumably the difference is due to the incompleteness of the list and the fact it is somewhat outdated. This implies that to identify the list part of an RDD sample, which is necessary for weighting, it is necessary to match the RDD sample with the list frame. Simply asking respondents whether they are in a directory is unsatisfactory. In order to avoid excessive matching costs and errors, it is important to establish simple and unambiguous association for the two frames, for example, matching only on telephone number and not by name or address.

I was surprised by the fact that a 98.2 percent telephone coverage rate is not considered high enough to permit the Canadian Labor Force Survey to rely completely on telephone surveys. In the U.S., that level of coverage is not even achieved with an area sample. If telephone samples seem to be satisfactory for all other purposes, it would be worth examining whether the noncoverage biases could be substantially reduced by differential weighting, somewhat similar to the use of weighting for nonresponse adjustment.

At Westat, such weighting is normally applied to RDD surveys to reduce the bias of the exclusion of non-telephone households.

The purpose of the dual-frame surveys in the papers presented here is mostly to increase the response rate although they also reduce the number of non-households called as compared to pure RDD. The price is additional complexity in the sampling operations and the possibility of error if the matching of the two frames is not done carefully, as well as the purchase cost of the commercial list. My instincts are that a more complex scheme should not be used unless there is a reasonably good pay-off. It is not worth taking the risk of errors for only minor gains. It would be useful to extend and repeat the experiments before reaching a final decision on conditions that warrant a dual frame designs.

We have occasionally used dual frame designs at Westat, but only to avoid very extensive screening for rare populations. In several health studies for a Government agency in which oversampling of the elderly was required, we obtained access to HCFA records and selected a sample of persons 65 years and over from the Medicare file and used RDD for persons under 65. For a study of the Jewish population in the Washington area, we obtained a list of known Jewish households from local sources and supplemented it with RDD. The dual frame aspects of the sampling did not create any particular problems in the survey operations. However, for the Medicare-RDD studies, we made sure the questions on age were very clear and specific. Similarly in

the Jewish population study, the sampling frames were defined in a way that simplified the matching. Dual-frame sampling considerably improved the efficiency of both acts of surveys.

4. Panel Surveys

Both the Canadian Labor Force Survey and the U.S. National Crime Survey are rotating, panel surveys. The Drew-Jaworski paper indicates that a telephone approach for the first time a household is in a panel appears to have a trivial effect on response rates in future periods. Census report showed substantial loss, mostly from difficulty in locating movers or from changed numbers. It's not clear why such a large difference should have occurred. It would be useful to find out what Statistics Canada did to keep alteration at such a low level.

In addition to effects on response rates, an RDD sample requires a somewhat different approach for sample rotation. The standard procedure, up to now, in the Census Bureau and Statistics Canada has been to view the panels as samples of addresses, and these are rotated into and out of the sample. I don't think this is feasible with RDD. The choice is to rotate the telephone numbers, or the persons. There are problems with both. If a panel consists of telephone numbers, then a sample of numbers that are nonworking or not residential needs to be included to guard against loss of movers. With a person sample, movers have to be located and followed up. It would be useful for both agencies to start exploring the effects of alternative methods of dealing with sample rotation.