

DISCUSSION

William D. Kalsbeek, University of North Carolina

The papers in this session all consider the problem of nonresponse but from slightly different perspectives and for varying types of survey investigations. They reflect the breadth of interest in nonresponse, which continues to be the proverbial Achille's heel of survey research. Four of the five papers presented here investigate attempts to reduce the effect of nonresponse by improving survey response rates. The fifth paper studies the implications of nonresponse on analysis in a population-based epidemiologic study.

Two general points emerged from my reading of these papers. First, one must take great care in extending the conclusions of these studies beyond the settings in which they were produced. A field protocol found to be effective in improving response rates in one survey might turn out to be quite ineffective in another survey where a different method of data collection is used, where the content of the survey questionnaire is altered, or where a different type of respondent completes the survey questionnaire. Second, it might be inferred from these papers that increasing the response rate and comparing respondents with nonrespondents are the only ways to accommodate survey nonresponse. This, however, is not the case since one might also measure the bias due to nonresponse or to create estimates which partially or completely remove it.

The first paper by Stopher and Sheskin suggests that response rates might be improved by completing data collection in two phases, the first a brief contact to establish a commitment to the survey by a large portion of the sample and the second a more extensive contact designed to cover the bulk of the survey questionnaire with as large a subset of the first phase respondents as possible. The idea of using mixed modes of data collection to increase response in surveys is not a new one. However, the idea presented here is novel in that the roles of the two phases are reversed from the usually mixed-mode approach in which a modest attempt at obtaining survey response is made in the first phase followed by a more intensive effort to convert nonrespondents to the first phase. As with the proposed approach, the intent of existing mixed-mode data collection schemes is to obtain the best response rate at the lowest survey cost.

Three local transportation surveys are presented to illustrate the feasibility of the approach being suggested. Two of the three survey illustrations appear to be tailor-made to exploit the advantages of this approach. In each of these surveys the initial contact was made in a captive setting where an extensive interview was impossible yet where personal contact by the interviewer was imperative. Response rates through the first phase approach 90 percent yet one might be concerned by the rather low (38 percent) response rate through the second phase since most of the important survey data are collected at that point.

One might speculate as to how high the response rates would be if the proposed approach were applied to the usual household survey. In this situation, it may be interesting, for example, to study whether the response rate after the second phase of the proposed approach would be higher than the response rate obtained by the usual

method of initial contact plus subsequent follow-up. The answer would depend to a large part on whether the first phase of the proposed approach would yield a higher response rate than the total efforts of the usual approach; that is, sufficiently higher to offset the nonresponse of the second phase.

The paper by Kerachsky and Mallar is an impressive effort to further assess the utility of monetary incentives, with a particular reference to longitudinal studies where the incentive payment is offered during periodic contacts with participants for whom financial remuneration for time spent in the survey is particularly meaningful. The authors suggest that a modest incentive of five dollars offered in this setting leads to a moderate increase in one's chances of reinterviewing participants and that the size of this improvement increases over time. Findings of the study also indicates that incentives may lead to improved data quality, although only indirect measures of quality were used. For the measures that were used, the effect of incentives on data quality appears to generally diminish over time. Of course a final decision on the feasibility of incentives in this setting will depend on whether the observed improvement in search efficacy and data quality would justify the added survey cost of the incentive payments.

The most troublesome aspect of the study is that notably greater search efficacy was indicated for the group receiving the incentive payment in the baseline survey even though the members of that group had apparently not been told about the payment until after they had been located. The explanation, as indicated by the authors, may be attributable to the randomization of the incentive/no-incentive treatment to the so-called comparison group. Measurement in the baseline survey was limited to this comparison group, who were those persons not enrolled in the Job Corps. The sample for the comparison group was chosen from seven area sites, each of which was randomly designated to receive the incentive or no incentive, and the treatment designated to a site was applied to all selected members of the comparison group. Because randomization of treatments was limited to a small number of sites, chance may have played the trick of having particularly adept field workers in sites where the incentive treatment was assigned. If this were true, one might also be somewhat concerned about the conclusions drawn from data in the first and second followup surveys where Job Corps and comparison group data were apparently combined. Specifically, one wonders if the incentive/no-incentive differences reported for the followup surveys are partially explained by the clustering effect in the comparison group. A simple way to test this hypothesis would be to exclude the comparison group from the analysis of followup data.

The paper by Ferrari and Bailey describes a study designed to compare the feasibility of telephone and personal-visit followup in dealing with nonrespondents to the 1980 U.S. Census. For the most part, the study was limited to a sample of single-household structures which had failed to return a census form on schedule to seven

district offices. For about half of the sample telephone numbers were obtained by matching addresses to address telephone directories (ATDs) available through the telephone company. Follow-up by personal visit was used for the other half of the sample. The sample was also evenly divided between structures sent the long form of the original Census questionnaire and structures sent the short form. Multi-household structures were excluded in all but one of the seven district offices because apartment numbers are not included in the ATDs, thus making matching more difficult. This immediately raises the questions as to why multi-household structures were excluded at all. One might speculate that perhaps the Bureau does not seriously consider telephone followup to be preferable to personal-visit followup in dealing with the multi-household structures, although this point is never stated.

Their preliminary analysis of the data from this study leads the authors to state a qualified preference for the telephone approach. Comparison of costs, item nonresponse rates, the percentage of marginally acceptable questionnaires, the proportion of duplicate administrations, and interviewer recruitment problems tend to favor telephone followup. On the otherhand, notably higher followup nonresponse and ATP coverage rates, which were less than 70 percent in four of seven district offices in the study, lead one to conclude that further consideration of the personal-visit approach to followup is warranted and that no clear-cut choice between the two competing approaches has as yet emerged.

Making a final choice between the telephone and personal-visit approaches will be difficult for a number of reasons. First, minor modifications in either approach may affect the recommendation that is made. For example, if the number of call attempts allowed for the personal-visit approach were reduced from four to two and if the number of allowed attempts for the telephone approach were increased from five to seven, the balance in a close contest between approaches might be tipped in favor of the telephone. Considering this point in light of the design for the study we are presently considering, it might have been useful to insert a few variations of both approaches to assist in the final decision-making. Of course, adding variations of these approaches might have made training more difficult. A second difficulty is how to anticipate the effect of time on our choice between approaches. Because of rising travel costs and the possibility of future improvement in ATD coverage, the practicality of the personal-visit approach relative to the telephone approach, may diminish rapidly between now and the next census. A later choice will be especially difficult if the ultimate findings of this study gives a slight edge to the personal-visit approach. Third, it is not clear how one might best choose between the two approaches. Certainly, the success rate in nonrespondent followup must be considered as should the quality of data gathered during the followup interview and the costs of each approach. The difficulty is in assigning the proper priority to each criterion since different priority schemes may well lead to different recommendations. (This dilemma, incidently, would also appear in the incentive study.) Fourth, the ultimate

recommendation from this study may not be a matter of a clear-cut choice between the two approaches. Results, instead, may indicate that both approaches should be used and applied to situations where they have been found to be preferable. For example, one might conclude that personal-visit followup would be best used in multi-household structures and in areas with low ATD coverage while telephone followup would be more useful elsewhere.

The paper by Kulka, et al. further extends prior research designed to fine-tune the data collection strategy used to maximize response rates in mail surveys. The present study is conducted as part of a national mail survey of registered nurses and is designed to measure the effect on response rates of such things as: the use of advance letters, personalization of survey stationery, the type of postage used in the original mailing, inclusion of a pencil in the survey packet, and the type of mailout used for the last followup attempt.

One is struck by the fact that differences in response rates between alternative procedures were seldom more than a few percentage points. Even the large variation response to followup that results from using certified mail or mailgrams implies relatively small variation in the overall response rate. The only notable exception, a substantially lower response rate reported when a personalized first-class envelope was used for the questionnaire, is probably due to an artifact in the survey protocol rather than one of the procedures being tested. Therefore, one can generally infer from these results that the size of the statistical nonresponse bias will also be virtually the same among the various procedures. Analyses of response patterns over time and the suggested measure of "response completeness" also suggest modest differences among procedures. The ultimate implication of this study seems to be that the best strategy in planning data collection for this type of mail survey is to pick the most reasonable but cheap combination of procedures rather than to search about for a more expensive combination which would at best yield a marginally higher response rate. In a way, it is reassuring to know in planning our survey protocol that we need not always rely on the more expensive device to produce the more impressive result.

The paper by Heilbrun, et al. differs from the others in that its primary objective is to assess the effect of survey nonresponse on analysis in a health-related survey by contrasting those who responded to a mailed questionnaire and were examined with those who were not examined prior to a ten-year prospective study of cancer mortality. The subjects of this study were a group of Japanese-born males living in Hawaii.

Several different measures are used in the respondent/nonrespondent comparisons. One of the measures, year of birth, is commonly used in respondent/nonrespondent comparison studies while the other measures of survivorship are somewhat unique in this type of study. Many of the differences computed according to these measures are small but nonetheless found to be statistically significant, due mainly to large sample sizes. For example, the largest statistically significant differences in the proportion

surviving (Figure 1) are less than 6 percentage points. The largest statistically significant differences are even smaller for the cumulative proportion who survive the risk of cancer (Figure 2) and the cumulative proportion who are alive, and free of other specific types of cancer (Figures 3-6). Larger respondent-nonrespondent differences are reported in the comparison of odds ratios of lung cancer calculated for smokers and nonsmokers.

In studies which investigate respondent-non-respondent differences it is important to note that the real effect of nonresponse is reflected in the difference between the estimate for respondents and the estimate for the combined

group of respondents and nonrespondents. The combined group represents the estimate produced when there is complete response and the respondent group represents the result of nonresponse on the original sample. Therefore, when reviewing the odds ratios of lung cancer (Table 2), the important comparison is not 19.1 versus 7.0 (i.e., examined versus unexamined) but 19.1 versus 15.1 (i.e., examined versus all men).

I would like to conclude my remarks by congratulating the chairman and the authors for an interesting group of papers on what continues to be one of the most visible and important topics in survey methods research.