DISCUSSION

A. L. Finkner, Research Triangle Institute

The two papers I have been asked to review are both interesting but widely different. The Gibson, Shapiro, Stanko and Murphy paper arose from an empirical examination of discrepancies in three well known statistical series. The Hartley-Biemer paper is theoretical in nature and extends results previously presented by Hartley and Rao.

With respect to the Census Bureau paper, it is good to see that its tradition of critical self-evaluation and public dissemination of the results is being continued. Although such practice sometimes makes an agency vulnerable to unwarranted attacks, it confirms the integrity of the agency and strengthens the entire statistical system.

The evidence presented by the authors certainly supports their general conclusions; i.e., vigorous methodological experiments which will give more precise estimates of the various components of total survey error are desirable. Such experiments are, of course, costly but as Jabine pointed out in discussing three papers at a recent Health Conference on Records and Statistics, more research on survey methodology is necessary if we are to get our money's worth on the vast sums spent annually on surveys. Perhaps the funds that have been recently appropriated for a Methods Test Panel at the Bureau of the Census will allow some such activities to be initiated.

In their first example, the authors contend that a questionnaire which was comprised of a supplement of attitude questions asked immediately prior to the regular victimization survey questions and which was administered to a random half-sample of respondents in 13 National Crime Survey cities, resulted in estimates of personal and property crime victimization rates that were significantly higher than those for respondents who were not given the attitude questions. Since the supplement considerably lengthened the interview time, there was some speculation that a possible fatigue factor might operate to give results in the opposite direction. Although the standard errors on which they presumably based their tests of significance are probably underestimated, the magnitude, direction and consistency of the estimated differences over all 13 cities leaves little doubt that the questionnaires with the supplement and those without it are measuring different things. The real question is which, if either, is closer to ground truth and how do we find out?

Their second example is not so clear cut. Although the questions concerning discouraged workers seem to have an influence, there does not appear to be any easy way of untangling the confounded effects. The statement is made, for example, that when discouraged worker questions are asked, there is a substantial increase in the estimate of the number of unemployed. Depending upon their definition of substantial, that statement is correct. If we accept the assumptions underlying Table 2, the estimates of unemployed were increased by 117,000 by asking the questions during months-in-sample 1 and 5 and by 75,000 when asking the questions in months 4 and 8. However, estimates of employment were increased 177,000 and 113,000 respectively. Apparently the questions stimulate an increase in both categories. Although the actual increases in employment estimates are greater, the percentage increases were, of course, greater for the unemployed.

Incidentally, the trends of month-in-sample estimates of total employment as published by Bailar in JASA in 1975 for the eight-year period 1968-1975 do not quite agree with the authors' graph. Bailar's data show a slight but real upturn in both months 4 and 8. The data in Table 3, relating to the period 1970-72 (T2), show the same general characteristics as those presented by Bailar while data from 1968-69 (T1) more nearly conform to the authors' graph. Perhaps adding the period of observation on the figure would clarify the point. The increase in months 4 and 8, shown by Bailar and (T2), may partially be the result of questions on discouraged workers.

The evidence in Table 3 seems to support the authors' contention more strongly. Consider the category of unemployed; if we subtract the index (T2) from the index (T1) for each month-in-sample, the differences should follow a particular pattern given the authors' hypothesis and no interaction between year and month-in-sample.

1. The differences in months 1 and 5 should be positive and substantial because the discouraged worker question is supposed to have an effect on (T1) but not on (T2). Those differences are 10.8 and 7.0 respectively.

2. The differences in months 4 and 8 should be negative and substantial since the discouraged worker question is postulated to affect (T2) but not (T1) in these months. These values are -8.4 and -7.2 respectively.

3. The differences in months 2, 3, 6 and 7 should represent estimates of differences between the two periods without the confounding effect of the discouraged worker questions. As such, they should be consistent and probably small. They are 1.2, -1.7, -0.2 and -1.5 respectively; probably none are significantly different from zero.

The data presented in Table 4 regarding the inquiry about acute conditions on the health interview survey are the most striking of any of the four examples. The fact that the differences appear to be independent of both age and sex is also worthy of note. Again it seems that a well designed survey experiment would be needed to ascertain the reasons for the discrepancies noted and to suggest corrective action. The information presented does raise some procedural questions. An acute condition is defined as one lasting less than three months. If it is first noted two weeks prior to interview, how will the interviewer know whether the condition will persist for more or less than three months? Also why should 68-75% of the reported conditions be those normally noted in the two weeks prior to interview? Even if the assumption is made that additional probing tended to define the date of onset more precisely, more than 55% of all acute conditions reported in 1973 and 1974 had their onset in the two-week reference period.

Tables 5 through 8 present another aspect of the phenomena regarding the supplemental questions concerning discouraged workers. Some of the entries in Table 7 and to a lesser extent in Table 8 are not self-evident; a description of how they were obtained would be helpful.

In summary this very interesting paper, as is often the case, raises more questions than it answers. It does rightly emphasize that if you contemplate adding supplemental questions to a continuing stable series, be prepared for some perturbations.

With respect to the second paper, it is encouraging to note that the important problem of measurement errors is being also attacked on the theoretical front. The paper to be published includes more material than could be presented at the meetings and addresses some of the questions that were raised at the session.

The authors point out that, in most current surveys, emphasis is placed on designs that yield precise estimates of target parameters while little attention is devoted to the precision of the variance estimates. They urge that more thought be given to the design of personnel allocation. From that standpoint, the added discussion on experimental (incomplete block) designs to be used as a basis for interviewer allocation is of particular interest.

The model which Hartley calls the mean square error decomposition model has been used in a number of practical situations and those results are available in the literature. The authors have indicated there is some interest in survey organizations trying the linear additive model; we look forward to results from its use.

There is a slight difference in the definition of $\varepsilon_{pst}$ and $e_{pst}$ between the material presented at the meeting and that submitted for publication. Although they both represent a combination of terms, the earlier version seems more consistent with subsequent development.