

DISCUSSION

Seymour Sudman, University of Illinois at Urbana-Champaign

The gestation period for the SIPP program is nearing an end, and there is eager anticipation for the birth that will occur in October. Although policy makers and economists are already licking their chops in anticipation of the findings, I would predict that there are uses of SIPP data that have not yet been anticipated, and that SIPP will become crucial as new government programs and avenues of research open in the decade ahead.

In their haste to get data collection started, policy makers have ignored the fact that not all the bugs have yet been worked out. This may have been the right decision. It is almost certainly the case that, like the other surveys conducted by the Census Bureau, particularly the Current Population Survey, revisions and refinements will need to be made as data are collected. While such revisions do cause difficulties, they are by no means impossible as indicated by changes on other federal data gathering programs, most recently, for example in the Consumer Price Index.

Kasprzyk's thoughtful paper, therefore, is not simply an abstract exercise, but effort to spell out a continuing program of research for the Census Bureau on SIPP that will ultimately enhance the reliability and validity and thus the usefulness of SIPP. I have no disagreements with any of Kasprzyk's list of issues. Let me briefly stress and expand on a few of them.

1. One can already anticipate substantial pressures between the need for timely data and the desire to have clean data. Financial data require substantial editing which could slow down the entire process. One wonders if the new technology may make it possible to do initial entry and editing on small portable computers that interviewers carry into the home to reduce editing delays and the need to call back households.

2. The discussion of sampling issues for SIPP is very complete. I agree with Kasprzyk that it is especially important to evaluate the use of dual frames of lists of program users and the area probability sample. As he points out the cost and difficulty of matching are greater than might at first be assumed.

I also agree that the costs of personal interviewing will certainly make it essential to consider lower cost alternatives. No discussion is included, however, of the possible use of mail or mail-telephone procedures in a dual frame approach or at least on some waves. For many respondents or for certain topics, it may be possible to get equally high quality responses by mail. At least, these approaches should not be omitted at this early date.

Turning to longitudinal issues, I am somewhat less concerned than Kasprzyk about the treatment of nonresponse. My lesser concern is because I don't anticipate any serious losses in cooperation across waves. Panel studies, such as the Consumer Expenditure Study, and those of commercial panel operators generally find only small dropoffs in cooperation after initial coopera-

tion, and the Census Bureau's previous record in obtaining high cooperation after interviewers are sufficiently trained should be an added positive factor. Item non-response will tend to decline slightly over time, if past experience is confirmed. Ferber found that in later waves of a panel study, respondents provided more information as their trust rose in the legitimacy of the requests for information.

The other two issues relating to changes in households and in the universe are not only technically, but also substantively interesting. A major issue is what happens to household or family income as a result of changes in family composition.

I must admit that my background makes me especially interested in the measurement related research, particularly on net worth. The easiest way to approach this research is to continue the analysis of the data collected by the 1979 panel. This can then lead to experiments for improving the forms and methods used in SIPP. Another priority item is to evaluate the quality of data from self and proxy reports. It will also be useful to see if there are differences in self-proxy quality for different household members.

As a final point, Kasprzyk discusses non-sampling error issues. As I have suggested, I'm not convinced in advance that non-response will be a serious problem, nor that interviewers will have serious impacts on cooperation, attrition or data quality. The other topic, recall bias, is an essential element for continuing research. It is not obvious that four month recall is ideal. Perhaps shorter or longer periods would be optimum taking cost as well as quality into account. Panel studies have other alternatives for improving recall such as the use of diaries or other forms of reminders such as records. Methods for making such procedures operational need development.

In sum, the users of income and financial data will become substantially better informed starting in October. It is to be hoped that data quality concerns will not be submerged in the excitement of putting this rich data source on line.

Turning now to the Richardson-Campbell paper, let me make a confession. I do my own income tax. I don't do it to save money, but rather as an intellectual challenge. Thus, I found the paper on improving IRS forms to be most interesting. One thing that doesn't come out in this paper that I looked for were estimates of the time or payments to others required to prepare a tax report. I find that it takes me several days to prepare my return, although most of the time is spent in gathering the proper records and checks, sorting into the necessary categories and checking the return to make sure that no deductions have been forgotten.

The IRS is charged with carrying out the complex laws that Congress has passed. Each complication is because of strong pressures from some interest groups to have their taxes re-

duced. Congress and the IRS need to have a good fix on the impact on tax return preparers of our complex tax laws.

This paper limits its discussion to the issue of improving the tax forms that are used. This is an important problem, particularly for large numbers of taxpayers who have relatively simple tax returns and who do not itemize deductions. For this reason initial research concentrated on the 1040A form.

The research described in this paper was initiated in late 1979 and consisted of the following steps:

1. Two rounds of focus groups
2. Field testing with volunteers using self-administered forms followed by self-administered and interviewer administered questionnaires.
3. A large scale test in Georgia (remember this was during the Carter administration.)

One can only applaud the use of research here and hope that this remains a continuing effort, rather than a one-time experiment. There are some questions about methods that were raised by the paper that I hope the author(s) can comment on:

A. The Focus Groups

The first question one has is who were the participants in the focus groups. Were these people who filled out their own tax forms or had someone else do it? What were their educational levels? Certainly, one does not look for a probability sample in focus groups, but one would want some of the participants to be the people who use the forms and have some difficulty with them.

B. Volunteers

The same question, of course, occurs about the discussion of the field experiment. How

were the volunteers selected? Were efforts made to include persons with low education levels? Other sampling questions include the size of the sample, and the administration of the experiment. Was it done by mail, at home, in malls, or how?

The paper indicates that time and accuracy data were gathered on performance. It would be nice to know the results. What were the time differences between the two forms? What were the levels of accuracy?

C. The Georgia Experiment

I had some difficulty in understanding the description here. Was it the case, as I concluded, that taxpayers were actually to use this new form for filing their return? If so, for what year? Was this a sample or the total population of Georgia? If a sample, how large was it?

Is there any information available on the fraction of recipients who actually used the forms they were sent? How many still had their returns prepared by someone else? How many used the old form although sent a new one? These statistics must also be considered in evaluating the new form. Also, if there are significant differences in return rates, these differences confound the error rates.

Was there information gathered on the time required to fill out the forms? Were there any time differences?

I recognize the severe time constraints that are always faced by conference presenters, but this paper needs more details in it to flesh it out. Let me repeat, however, that the basic idea is an excellent one--forms should not be invented by lawyers, but by form specialists and then tested on real-world populations.