

DISCUSSION

Joseph Waksberg, Westat, Inc.
1650 Research Boulevard, Rockville, Maryland 20850

I would like to start by complimenting the authors and the organizer of this session for the comprehensive analyses of what is, I think, a quite important research study. Federal agencies have a habit of getting absorbed in the design and operation of methodological research but running out of money just when the data collection ends, and then short changing the analyses. It is refreshing to see a detailed examination of the experiment, including having the same data viewed in different ways. More important, however, is the fact that we have research that is not only important for the Consumer Expenditure Survey, but that is also applicable to a wide variety of surveys in which diaries are required or encouraged.

I plan to briefly summarize the main features of the analysis plans, comment on the principal conclusions and suggest some changes in the analytic approach and directions for future research.

1. Review of Analytic Approaches in the Papers

Greg Sliwa's paper contains comparisons among the three experimental diary formats and with the production sample. The data that are compared consist of mean weekly expenditure per consumer unit for major classes of items, number of expenditures reported for each of these classes, and the percent of consumer units reporting each class. Differences between pairs of procedures are calculated and tested for statistical significance.

Tucker's and Bennett's paper repeats some of the same analyses as Sliwa, but carries out tests of significance for groups of items which take into account the joint probabilities of the members of the group. Tucker also reports on other outcome variables -- respondents' attitudes and behavior, expenditures in first day and first week of reporting period compared to later time periods, response rates, and the proportion of diaries requiring recall interviews.

The paper by Vitrano and co-authors also examines mean weekly expenditures, but focuses on a subset of items that are specifically referred to in one or another of the diary versions. The extent to which such cues affect reporting on the diary is examined. This paper also reviews the time period effect and the extent to which respondents actually filled out the diaries as compared to recalling expenditures at a later point in time.

The fourth paper looks at a different aspect of the quality of the data, the extent to which central office editing indicates reporting problems.

2. Issues the Papers Do Not Consider

It is useful to recognize issues the papers do not consider. I would like to call attention to three subjects that might be taken up in future papers.

- (1) The papers do not compare total expenditures for classes of items with independent data as was done by Bob Pearl about 10 years ago in his analysis of the quality of the 1972-73 Consumer Expenditure Survey. The three papers that examine the total expenditures implicitly take for granted that more is better. Pearl's reports did indicate that household reporting of almost all expenditures tended to understate the true levels of expenditures, and almost all other research on household expenditures is consistent with this. Consequently, the assumption that diary formats which

produce higher levels for most items are better than those with lower values is a reasonable one. However, a comparison with independent data might shed light on the possibility that some of the differences are due to misclassification errors rather than omissions. As some of the authors have pointed out, there is at present only indirect evidence on the sources of difference.

- (2) A second issue not taken up is the extent to which an improved diary format would influence the CPI. The effort and cost of the Consumer Expenditure Survey is justified by its use in the construction of the weights applied to price changes for individual items measured in the CPI. For this purpose, the proportions of the expenditures among the various items are relevant, rather than the level of expenditures. Tucker does calculate these proportions, which he refers to as the relative importance of an item, denoted by RIM. There are small but measurable differences among diary types in the RIM. It would be interesting to see how much effect these differences have on the CPI.
- (3) Finally, the authors do not make any recommendations on whether, or how, the diary on the ongoing CE survey should be changed. Some of the implications of the research appear clear. The specific diary format stood out as preferable in all four papers. However, it is not clear whether the current diary should simply be replaced by the specific diary, or there are some features of the other diaries that should be attached to the specific diary format.

3. Discussion of Experimental Results

3.1 Expenditure Levels

It was pointed out by Sliwa that most previous research in which diary formats were compared showed higher amounts of reporting for diaries with formats organized and labeled by product category. Consequently, it is not surprising that the specific diary showed higher expenditures than the other two diary versions. This pattern was clear in the three papers reporting on expenditures, although the significance levels were not always the same.

I think several papers have probably understated the extent to which differences between pairs of procedures exceed their standard errors. The three papers take somewhat different approaches to confidence limits. The paper presented by Vitrano estimates a design effect of 1.5 from past data and applies it to simple random sampling variances. Tucker uses BRR and assumes independence among the samples, but takes into account the joint probabilities of differences for groups of items. Silwa is apparently the only one who actually calculated the variances of differences between procedures.

The three samples were not independent since they were deliberately interpenetrated within PSU's, segments, and within interviewers. This undoubtedly introduced positive covariances among the samples, which would result in the standard errors being smaller than the ones computed on the assumption of independence. There is no information on how much the standard errors are exaggerated, but I think that instead of having a design effect of 1.5, as assumed by one author, it is likely to be less than 1. Proper tests of significance would probably show that many more

differences between procedures are significant than are indicated by the authors.

With BRR, it really is not difficult to take the covariances into account in estimating the standard errors of the differences. Silwa did this by calculating the differences between pairs of procedures for each replicate separately, and computing the variances of these differences among replicates. Some additional work might be necessary in adapting these result to the T^2 measure that Tucker uses, but it should be possible to develop some reasonable approximations to this.

The comparisons can also be refined in another way. Household's expenditures are affected by a various demographic factors, such as family size, income, geographic location, etc. Expenditures for these domains were calculated but the sample sizes were too small for much meaningful analysis of the domains. Variations in these factors among the experimental groups could increase the variances of differences between procedures. This additional component of variance could be reduced by standardization of the samples, or by using multivariate techniques for analysis.

I have one other comment on the comparisons of average weekly expenditures. As I mentioned earlier, several papers pointed out that differences in reporting levels could be due to omissions or to misclassifications. The paper by Vitrano and his colleagues discusses this issue at some length, but there is only indirect evidence of misclassification and it is not clear whether or not this is a major problem. For the construction of weights in the CPI, misclassification is probably as serious as omissions. I suggest that more attention be paid to distinguishing between sources of errors. Earlier in these comments I noted that comparisons with independent data might indicate problems of misclassifications. Another possibility would be a follow-up, probing reinterview in which respondents are asked to describe some of their diary entries in greater detail.

3.2 Response Rates and Recall vs. Diary Entries

The three experimental diary versions all had about the same response rates and they were about 10 percent lower than in the production sample. The authors attribute the difference to the greater experience of interviewers on the production samples. One of the goals of the experimental diary formats was to make them more attractive to the respondent and less forbidding, and when I examined the diaries I thought they accomplished this goal quite well. However, this did not appear to affect the response rates. Could it be that such efforts do not matter?

The extent to which respondents actually fill out the diaries is related to the issue of nonresponse. Whether households that accept diaries but fail to fill them out should be considered as responses or nonresponse is, to some extent, a definitional matter. The three experimental diaries had about the same proportion of cases that were completely filled by the respondent, 71 to 75 percent. However, these values were about 10 percent higher than is currently achieved by the actual Consumer Expenditure Survey. One possible reason is that the improved appearance and size of the experimental diaries encourage continued cooperation. There is another possibility. If the higher response rate for the production diary was accomplished by getting reluctant household members to accept diaries, it may be these types of respondents are not very conscientious in filling them out. The proportion of all sample households, respondents and nonresponses, that completely filled out the diaries was about the same for the CE and the experimental samples.

If what occurs in the CE is that reluctant households are persuaded to accept diaries, and then fail to fill them out, then it becomes dubious whether the higher response rates

really produce better quality data. I am not sure that recall of expenditures after a week or two has passed is better than imputing data for these cases. I think it would be useful to explore this issue to shed light on what would be a sensible strategy for reluctant cooperators.

3.3 Attrition in Reporting Over Time

The authors are rather cautious in saying there was no statistically significant evidence of reduction in expenditures reported in week 2 as compared to week 1, although expenditures in day 1 of week 1 are clearly higher than the other 13 days. The weekly differences follow almost exactly the pattern observed in both the 1980-81 and the 1972-73 surveys. The differences struck me as quite large, with the second week about 8 to 15 percent below the first week for the three methods. I am surprised they were within the bounds of sampling error. The overestimates of sampling errors of differences discussed earlier probably apply to this analysis also. In fact, they may have even a greater impact, since in comparing weeks the covariances within households of expenditures for the two weeks should be taken into account.

The changes in diary formats do not seem to have improved the second week's performance relative to the first. Also, the fact that the higher level in the first week is primarily due to the first day is also consistent with earlier experience. A high priority subject for future investigation should be to explore reasons for the first day phenomenon.

3.4 Respondent Typology

The analyses of respondent typology reported by Tucker seems to indicate that respondents reacted about the same to the three experimental diaries. Typology thus does not provide a basis for selecting the preferable diary. It would probably be useful to review the results to see if they indicate necessary changes in the interviewers' training or their approach to respondents.

It is somewhat ironic that all of the attention given to make diaries attractive and appealing to respondents does not seem to affect their attitudes toward the survey. In fact, if one looks at the response rates, it appears that building up an experienced interviewer staff is much more effective than worrying about diary appearances. This should be an important factor in considering the desirability of a continuous CE or one carried out at fixed time intervals.

3.5 Data Adjustment

The analysis of data adjustment required for the three diaries indicates that there were fewer problems with the specific diary than the other two. It is nice that different criteria all point to the specific diary as the preferable one. However, none of the error rates is very large so that from the viewpoint of data adjustment, all of the diary versions seem to be acceptable.

It is interesting that the control diary had a much higher proportion of items with missing costs than the production diary. I assume the more experienced interviewers who worked on the production diary did a better job of editing when they picked up the diaries. Perhaps new interviewers need more training on how to review the diaries and encourage respondents to supply missing information.

4. Direction of Future Research

About 20 years ago, in his report on Methodology of Consumer Expenditures Surveys, Bob Pearl described what he considered high priority subjects for research for the diary component of the CES. The format and organization of the

diaries was the highest priority item on his list. A second item was whether monetary incentives would increase the response rates significantly. The Bureau of the Census carried out such a study in connection with the 1972-73 CES. This still leaves the following five subject for investigation:

- Length of time respondents are willing to complete diaries;
- Special recording procedures when documentation of purchases are available, e.g., credit card purchases, cash register tapes for food purchases, etc.;

- Which items are best suited for the quarterly survey interview and which should be on the diary;
- Whether conditioning takes place during diary keeping so that respondents change their buying patterns; and
- Whether more household members should be involved in the diary keeping.

I look forward to being at future ASA meetings discussing results of research on these subjects.