

REPORT ON BLS CONFERENCE, JANUARY 1987
Colm O'Muircheartaigh
London School of Economics and Political Science

This summary of recommendations for redesigning the U.S. Consumer Expenditure Survey (interview and diary portions) is presented in two sections. The first section outlines five major categories of issues discussed at the conference, each of which, though related to the others, suggests ways in which the general problems of the surveys may be tackled. The second section examines problems with the current survey design, and presents specific recommendations for designing a research program to serve as the foundation of the redesign effort. Top priorities for this research are programs to improve our understanding of the survey response process, and programs to identify and select appropriate respondents.

I. BASIC ISSUES

1. Underlying the whole discussion is the issue of objective measures of quality for the output from the surveys. This requires the development of objective criteria for evaluating the overall quality of the survey data-- criteria which take into account the three basic features of survey quality: unit nonresponse, item nonresponse, and response errors.

Looking back on our assessment of problems with the current surveys, it is surprising how often our identification of problems was based on some combination of (1) personal experience as respondents to the CE interview (gleaned through the mock interviews that were arranged prior to the meeting) and (2) the application of our particular expertise in survey research and/or cognitive psychology in providing professional assessments of the survey instruments. Without discounting the value of our collective personal experiences and professional judgements, we should acknowledge that our identification of problems with the survey was largely impressionistic. (This may be the reason why the bulk of the research recommendations which we discussed were designed, to a large extent, to confirm the existence and determine the exact nature and source of existing problems with the survey data.) In any event, it is clear from our discussions that the development of measures of quality and the documentation of existing problems are high priority items for the CE redesign research program.

2. In order to evaluate the quality of the survey data it is also necessary to formulate evaluation criteria for each section of the instrument (e.g., category of expenditure). The requirements of the data users in terms of completeness of information and precision of responses must be known in order to make it possible to establish priorities for sections of the questionnaires and for items within the sections.

A top priority for the CE redesign research program is a clarification of the goals and purposes of the surveys and the requirements for the data. BLS must, in consultation with the data user community, the Census Bureau, etc., answer such difficult (but basic) questions as: how accurate must responses to

specific items or segments of the survey be to meet the analytic needs of the survey? Is accuracy more important for some items than others? If the effort required to secure accurate responses to selected items increases burden to the extent that there is an increase in unit nonresponse, is it possible to retain high response rates by eliminating other items, or turning to alternative data collection methodologies for some items? With respect to the recall of major purchases, is more accurate price information more or less important than accurate recall of the time of purchase? Is it necessary for price or timing information to be precise, or is information provided within specified ranges sufficient? Would failure to report accurately the precise timing or amount of a purchase pose a greater problem than failure to recall an expenditure entirely?

Raising these issues, which are CE-specific, essentially implies the need to decide whether bad data are better than no data. If various design choices result in tradeoffs among unit nonresponse, item nonresponse and the accuracy of response to specific items, which of the three is to be valued more highly than others? The answers to such questions provide the only meaningful criteria for choosing among design options. Consider the selection of appropriate respondent rules. If only very small inaccuracies in the data can be tolerated, the "decision-rule" for formulating "appropriate" respondent rules will probably sacrifice response for quality data. If less accurate data is acceptable for the purposes of the survey, however, proxies (therefore, lax respondent rules) may be more tolerable--(sacrificing quality for response). It is impossible to maximize the overall quality of the data if the actual goals of the survey and analytic needs for the data are not identified.

3. The utility of the information generated by a survey is directly related to the level of respondent participation. There are many possible levels of participation, ranging from unit nonresponse (failure to participate) at one extreme to complete and perfectly accurate response at the other. Most individuals will fall between the two extremes and the design decisions taken will determine where in the continuum the individual will be located. While we currently do not understand the exact nature of the tradeoffs which may occur among the three major response problems (unit nonresponse, item nonresponse, and data quality/accuracy of individual responses) when survey designs are modified, it is reasonable to assume that such tradeoffs occur. Thus discrete improvements in response rates or the quality of responses (in general, or for certain segments/items on the survey) may be accompanied by net decreases in the overall quality of the survey data. For this reason neither specific response problems (e.g., nonresponse, accuracy) nor the impacts of design modifications should be considered in isolation.

Similarly, it is important not to consider the cognitive and motivational bases of survey response in isolation from each other. One could argue that motivation of the potential respondent underlies all data collection procedures. At a minimum, the individual must be motivated to agree to participate in the survey. If individual respondents are able, but unwilling, to provide answers to our questions, the overall quality of our sur-

vey data suffers. On the other hand, one could argue that ability to perform the task as it is presented is a prerequisite for obtaining high quality information. If some respondents are willing, but unable, to provide the requested information (e.g., because they are unable to comprehend and undertake the task as we present it to them, because they do not have access to the information), once again the quality of the data is imperilled. Both components are critically important in determining data quality overall.

The implications of these observations for BLS are straightforward. First, they underscore the need for developing objective measures of overall data quality (see item 1, above) which take into account all three of the basic features of survey quality: unit nonresponse, item nonresponse and response errors. Second, they remind us of the need for interpreting our discussion, and subsequent recommendations for a research program, in a wider context. Most of our discussion focussed on ways in which elements of the current survey designs imperil data quality by providing respondents with tasks they are unable to perform. The motivational (willingness-to-perform) factors received much less attention. Clearly it is reasonable to restrict the focus of attention at a two-day conference. But unless research demonstrates that cognitive factors pose greater barriers to the overall quality of response than motivational factors, it would be unwise to continue to focus attention (and resources) solely on the cognitive aspects of the survey response process.

4. Most of our discussion and research recommendations focussed on the role of the respondent in the survey response process. We should not lose sight of the fact that the interviewer is the critical intermediary between the survey design and the respondent. As such s/he constitutes a source of error but also provides a source of information on the degree to which the survey tasks are being implemented properly. In designing a research program to guide the CE redesign effort, BLS should examine interviewers as sources of error, should consider how interviewer/respondent interactions may be used to motivate fuller and more accurate responses, and should tap the interviewers as a source of information on problems with the current survey design. It is probably safe to assume that interviewers and their supervisors have far more experience with respondents' problems with the surveys than do most of the survey administrators and the methodologists/researchers associated with the surveys. Even something as informal as debriefings of key Census Bureau field staff (e.g., at the regional meetings of the senior field representatives) would be an inexpensive means of gathering valuable qualitative information.

5. The task itself is the fifth major component. The content of the task is determined by the criteria (objectives) mentioned in item 2 above. The organization and execution of the task currently is determined implicitly by the questionnaire layout and the interviewers' instructions. The extent to which the potential respondent is informed of the nature and magnitude of the task and the point at which this information is communicated constitutes a major element in the survey design. Deci-

sions must be taken as to the respondents' input to the organization and execution of the task. The fact that the interview is repeated (possibly with the same respondent) on four further occasions should be taken into account in these decisions.

These five categories of issues constitute the core of the CE redesign research program. In my opinion, BLS should give top priority to developing objective measures of quality for the survey output, and formulating evaluation criteria for each section of the survey (i.e., completely reevaluating the goals and purposes of the survey, clarifying research priorities and considering, where appropriate, the feasibility and desirability of implementing alternative data collection strategies). Throughout the redesign process, BLS should examine tradeoffs between response and quality which may be generated by various design modifications. The research effort should not overlook the role of motivational factors in influencing response quality. It should examine interviewers as a source of error, and as a source of information about error. Finally, to the greatest extent possible, BLS should build flexibility into the redesigned survey program--developing questions and questioning strategies which allow for wide variations in how respondents approach and respond to our requests for information.

II. PROBLEMS WITH THE CE

The three elements of the current CE (interview) survey design which seem most important in terms of poor data quality are: (1) length and complexity of the survey instrument; (2) length of the recall period; and (3) respondent rules. Specifically, it appears that

- (1) the current survey instrument is too long and complex (leading interviewers to rush through items and losing the respondents' interest);
- (2) the recall period is too long for many respondents (in that they are unable, either in an absolute sense, or in the context of the current survey's bounding and memory cues, to recall accurately the occurrence or timing of certain expenditures);
- (3) the current respondent rules are too lax (failing to identify and/or require that the interview be conducted with the "best," most knowledgeable respondent -- i.e., valuing high unit and item response rates above accurate responses).

With respect to the diary portion of the CE survey, the two elements of the current survey design which appear to be responsible for the bulk of the data quality problems are: (1) respondent rules and (2) questionnaire design. Here, too, it seems that the current respondent rules are too lax, and that the existing survey instrument suffers from poor (often ambiguous, at other times merely difficult-to-understand) question wordings.

While we generally agreed that these elements of the current sur-

vey designs contribute significantly to data quality problems with the surveys, there was speculation, but little agreement, as to what, precisely, the exact nature and scope of these problems are. It is therefore not surprising that the bulk of our recommendations call for research to identify the scope and source of existing data quality problems. Somewhat surprisingly, many of these recommendations, while framed in terms of the CE content, require more general investigations of the cognitive aspects of the survey response process--most notably, how individuals store, organize and retrieve information. The major advantage of framing this research in terms of the CE surveys is that, in addition to learning more about cognitive processes associated with survey response generally, specific problems associated with these particular surveys (e.g., ambiguously worded items) will be identified as well.

Two research areas to which we devoted considerable attention were: (1) research which is intended to improve our understanding of the survey response process (in general, and with reference to the CE interview- and CE diary-surveys specifically); and (2) research to identify and select appropriate respondents. A third area for research which received some (perhaps insufficient) attention was design modifications intended to affect respondent motivation. Some specific recommendations for research in these areas which I feel merit further consideration are described below.

A. Research which improves our understanding of the survey response process

1. General Aim

The task of designing surveys to elicit accurate information from respondents is hampered by an absence of information about the ways in which respondents store, retrieve, and organize information. Developing a better understanding of these important cognitive components of the survey response process should help the researcher to frame requests for information in a way that is easier/possible for the respondent to understand. We assumed that respondents who are currently unable to provide the requested information (or unable to provide it well/accurately) can be rendered capable of complying with our request if we pose questions in the right way, or provide the right cues/prompts to enable the respondent to recall and present the information in the desired format. The ultimate goal of such research is to enable the survey researcher to develop questions and questioning strategies that reflect respondents' orientations and assist them in formulating their responses.

2. CE Interview

The following issues are of particular interest:

What initiates the recall process? (important to know because it will tell the researcher the best way to cue recall of purchases)

Are respondents more likely to remember

purchases when they think first of events associated with the purchase, or when they make mental lists of possible purchases (eliminating some and recognizing others)? How important are bounding events, such as holidays and trips?

Is it best to bound the referent period by using calendars or time lines for respondents to refer to visually?

Is the bounding function improved when respondents are encouraged to frame the time period in terms of events of importance to them (e.g., personal holidays)? (Is using calendars or key events to fulfill the bounding function more useful for dating purchases accurately, or for cueing recall of -- particularly unusual -- purchasing associated with major events?)

Is it possible for respondents to recall accurately all the aspects of a purchase asked about in the survey (e.g., time of purchase, amount of expenditure) without referring to records/receipts?

What is the magnitude of error associated with the recall of purchase amounts?

Is there a consistent pattern of overreporting or underreporting for some items? To what extent would it be worthwhile to collect only data on amounts of items purchased and to use other sources to collect data on costs?

A number of possible approaches to these issues were suggested:

The diary sample could provide a good vehicle for recall experiments, though it would be desirable to design an experimental diary which included more of the items on the interview schedule. A subset of those who had completed the diary could be visited on later occasions--one week, one month, and three months, for example--and the same information collected by recall. The discrepancies would provide a lower bound for the estimate of the magnitude of error.

The time use literature provides a means of learning more about the best way to cue recall.

The findings in cognitive psychology on part-whole cueing are relevant to the design both of the questionnaire and of the diary. A systematic investigation of the impact on reporting of short and long lists of examples for different types of expenditures is required. It is not obvious that

the same strategy would be optimal for all classes of expenditure.

The possibility of using recognition rather than recall to cue reporting was advocated by Mingay-- for example by asking respondents to look through their wardrobes in order to answer questions on purchases of apparel.

In the case of proxy reporting, the suggestion was made that reporting might be requested for each person separately for the whole questionnaire rather than section by section for the whole consumer unit. This would of course add considerably to the length of the interview. An experimental comparison of the two procedures would be necessary to evaluate this strategy.

One major opportunity exists in the current procedure to estimate data quality in the CE interview. Information is collected for the month of interview as well as for the three month referent period. The subsequent interviews should record whether the specific expenditures reported as having occurred in the month of interview are also reported three months later and the dates for which they are reported. Because of the variation in date of interview this would not provide ideal validation, but it would provide a basis for direct comparison of short-term and long-term recall for these items. It would also provide some indication of the types of respondents and the types of purchases that are particularly prone to problems of omission and telescoping. The cost of this modification would be low and I would recommend it strongly.

There is a development of this idea which would be of considerably greater benefit; if four months' data were collected each quarter, there would be one full months' overlap between successive interviews and the comparison of the two sets of reports for that month would provide an excellent basis for the assessment of telescoping and omission. An alternative would be to collect three months' data as before but to have interviews at two month rather than three month intervals.

This recommendation may be too drastic to allow its incorporation in the survey directly. It seems to me, however, that unless such an investigation is carried out, speculation and prejudice will continue to provide the basis for the assessment of the magnitude and type of errors. It may well be desirable to execute a parallel experimental survey design in order to assess the impact of such changes before modifying the main survey in such a fundamental way.

3. CE Diary

The recall period for the CE diary is assumed to generally be no longer than 24 hours, though interviewers may record purchases made over the entire course of the record-keeping period through the respondents' recall when the diary is picked-up, if items have not been filled-in previously. For this reason, recall re-

search emerges as a lower priority for the diary than the interview portion of CE. Nevertheless, a better understanding of how information is stored and retrieved may aid us in improving the diary as well--particularly with respect to "unremarkable" (sometimes small, sometimes routine, for whatever reason easy-to-forget) purchases. It was suggested that it might be possible to cue respondents to recall purchases by reviewing the day's activities and/or providing lists of easily-forgotten activities that may be associated with purchases (e.g., lunch time snack at office cafeteria). Providing such cues after the space for initial recording (to capture "forgotten" purchases) and instructing the respondents to report such purchases in a different location would have the added benefit of enabling the researcher to identify purchases or classes of purchases which are most frequently overlooked.

In addition to benefiting from knowledge about the recall process, the CE diary redesign effort would benefit from knowing how respondents categorize information (including the basis for their decisions to classify purchases in the categories used in the experimental diaries). Tanur made a suggestion that would be valuable both for learning more about this classification process and the recall process. She suggests asking respondents to keep diaries using completely blank sheets of paper. Presumably one would be able to infer from the way in which purchases are recorded (e.g., grouped according to time of activity/purchase, amount of purchase, who goods were purchased for, type of item purchased) something about how respondents remember and "naturally" group the information.

Though this may be useful, there are two potential drawbacks: (1) respondents may make notes not included on the form, "organizing" the information more attractively, etc., for the actual diary, thus obscuring their "actual" thought processes; and (2) respondents may go through much of the "grouping" and recall mentally, only writing the information down once it has been remembered and subsequently reorganized in a more "attractive" or "coherent" fashion.

Progress toward evaluating the usefulness of the diary could be made at little or no cost if interviewers were instructed to flag those items that are filled-in on recall (versus those recorded by the respondent before the interviewer arrives to pick-up the diary) as a matter of course. This would permit some analyses of the differences between the two sorts of data, even in the absence of validating information.

4. Possible Methodologies

While there are many means of learning how respondents actually approach the survey response process, the following four are particularly appealing, as they can be conducted on a relatively small scale in laboratories rather than in the field: (1) "think-aloud" interviews; (2) behavior-coding pretest interviews; (3) observations of respondents filling-out questionnaires; and (4) retrospective protocols. Two of these methodologies, the behavior-coding interviews and the retrospective protocols, can be conducted in tandem with "live" field work as an alternative

which is particularly desirable if one has reason to believe that the types of individuals who are easily recruited for or are willing to participate in laboratory studies are unrepresentative of the population.

Think-aloud interviews can be used both to discover how respondents formulate responses to survey items, and to uncover specific problems with individual survey items (e.g., ambiguous wordings). In the course of these interviews, respondents are encouraged to report thoughts, perception, and understanding/misunderstandings. The interviews are conducted to discover how respondents interpret survey questions, retrieve information from memory, categorize events and expenditures--in short, how they arrive at their answers to survey questions. Given the length of the CE interview, many participants suggested breaking the interview into sections, and conducting the "think alouds" for only one or two sections of the interview with any given respondent (to avoid fatigue). As this is exploratory, hypothesis-generating research, the number of participants recommended is relatively small--(Mingay says 50-60 subjects per group would be adequate; Bishop recommends 200-400; Schaeffer says small, geographically concentrated but heterogeneous samples of 50-100 respondents should suffice).

Behavior coding pretest interviews indicate at which points throughout the interview the respondent has difficulties with the survey instrument (and, most often by inference, what the nature of the problem is). Simultaneously, they indicate at what stage(s) the interviewer is having difficulties, and can be used to monitor interviewer performance. The technique is based on identifying behaviors which indicate that either the respondent or the interviewer is having some problems with an item (e.g., respondents asking what a word means, or giving an inadequate or inappropriate answer, suggesting failure to understand an item, or inability to remember an event/expenditure). Such behaviors are coded to yield a frequency of behavior types per item; the frequency and variation in behavior types may be used to identify problem items and in some cases to identify the nature of the problem.

Another way of noting items which cause respondents problems is to observe respondents filling-out forms. It also may be possible to infer how the respondent formulates answers to specific items (e.g., when asked to report types and amounts of expenditures over time, does s/he write down first what was purchased, second when it was purchased, and third what it cost, or are these pieces of information recorded in some other order). It seems to me that there is a real problem with

this--unless respondents write down everything they think of, exactly in the order they think of it one could make incorrect inferences. I also question how useful this technique would be for identifying problem items. Should one infer that respondents that take more time on items are having more problems than those that don't? Maybe they are just careful respondents; or maybe some items require more thought than others. Personally, I like to keep inferences to a minimum; I agree respondents may not know what they do or why they do it, or may be reluctant to report difficulties they have understanding things due to social pressures, but it still seems to me that verbal reports are less prone to error than inferences from observed behavior.

The other commonly recommended method for learning more about respondents' survey-response processes is to administer retrospective protocols after the respondent has participated in the survey. These post-survey surveys are used to uncover problems which the respondents encountered in answering the survey. They also may be used to question the respondent about how s/he formulated responses to particular items (thereby gleaning insights into information storage, organization and retrieval processes). A third, related use is to ask the respondent how certain s/he is that the information provided is complete and accurate.

The importance of the interviewer in the data collection process is worth reiterating here. Experienced interviewers will be able to provide a good deal of information on the success or otherwise of different questioning procedures, questions and respondents. In a study I carried out for the World Fertility Survey, for example, interviewers in three different countries were asked to classify respondents as very good, good, fair or poor in terms of reliability and level of cooperation. In all cases there was a strong correlation between interviewers' ratings and the simple response variance of the responses. It might be worth while including a number of interviewers and/or SFRs in any future conference on data quality for a particular survey.

B. Research to identify and select appropriate respondents

The major impediments to data quality in surveys are comprehension problems (stemming e.g., from ambiguously or poorly-worded items) and inability to recall events/activities accurately (a problem which may be exacerbated by failure to develop questions and questioning strategies in ways that reflect respondents' orientations). The belief that better questioning strategies and better-written items will improve data quality rests, in large part, on the assumption that the respondent has access to accurate information--the challenge is to help him/her access and report it.

Whether or not this assumption holds for respondents who are reporting for themselves, it is almost certainly not the case for

proxy respondents. In the most extreme case, proxy respondents are unable to report not just high quality, but any, information for other members of their consumer units or households when they simply don't have access to the information (i.e., never knew the answer to the question). Equally problematic are those cases in which the respondent has some information to report, but not complete or accurate information. (In the absence of perfect communication, it seems unlikely that one member of a consumer unit would know every purchase made by every other member of the household--what was purchased, when, and for what sum of money.) When this inability is represented by item nonresponse, both researcher and end-user of the data have concrete evidence that something is missing--it helps to remind them of the limits of the methodology and the data. When an answer is given, however, it may be too easy (particularly for end users of the data) to forget that it is not necessarily the right answer.

The best strategy for dealing with the potential data quality problems posed by the use of proxy respondents is twofold. First, we need to identify the "best" respondent(s)--the individual or individuals within the consumer unit who have access to the information being sought by the interviewer. Second, we need to determine the effects on data quality of using alternate respondents. This second piece of information is critically important if we are to develop a reasonable set of respondent rules--one in which the identification and selection for interview of "appropriate" respondent(s) is based on a considered, fully-informed decision regarding the implications for data quality, and the tradeoffs which may have to be made between securing responses and obtaining complete and accurate information.

The best respondent for a particular question or series of questions will depend on the nature of the household and on the type of expenditure. There are two potential sources of information on this issue.

First, it should be possible by analysing existing data to compare reports obtained from different types of respondents, perhaps using the classification of relationship to head of household. If the information as to which responses are self-reports and which are proxy-reports is not currently coded this is clearly a first and low-cost innovation which would make such an analysis possible. If the analysis reveals substantial differences in levels of reporting, this would indicate the need for further targetted research in the area. (It is of course necessary to bear in mind that differences in level for a non-randomised comparison of this kind would not necessarily imply differences in data quality but would suggest the desirability of further investigation). It is likely that no simple category of respondent would be best for all types of expenditure.

In addition to the analysis of existing data, some field experiments are needed to establish the effects of the present field strategy. A field experiment should be carried out to compare two procedures: (i) the standard CES procedure; and (ii) separate interviews with all members of the consumer unit. The comparison of the results would provide an indication of the

differences in costs, in level of reporting, and in nonresponse for the two procedures. The reactions of the field staff and field directors to the experimental strategy would also constitute an important element in the results. The experimental group would not be a part of the regular CE sample, but might be matched with a subset of the CE sample; in other words the costs of the experiment would be confined to the cost of the additional sample receiving the experimental treatment.

Second, in parallel with the investigations suggested above, research should be carried out to identify the "best" respondent(s). Such respondents need to satisfy two criteria: they need to have access to the information being requested and they need to have the ability to report it adequately. The size and composition of the household will be critical in determining which, if any, individual can report for the whole household.

In this context we should note that it is unrealistic to expect that there will be a single answer to this problem. It is necessary to identify kinds of purchases which will typically be known only to the purchaser; it is also necessary to determine who in the household will have sufficient information on other expenditures. This implies the need for the construction of a typology of expenditures based on the access of different potential respondents to the information being sought. Such research in its initial stages will require intensive interviews in a small sample of multi-person households; the existing knowledge of interviewers and SFR's would also be invaluable in planning and executing this phase of the research.

The next stage would be to investigate the practicality of having the interviewers choose the most appropriate respondent in the field. This would involve the preparation of a set of questions and decision rules for the interviewer to apply in each household/consumer unit, and a field test of these procedures.

C. Design modifications intended to effect respondent motivation

Participants paid relatively more attention to cognitive than motivational bases of data quality problems with the CE interview and diary. This was particularly evident in the case of recommendations for research to understand more about the survey process; overall there was more concern about understanding the cognitive aspects of the process than with determining the importance of motivational factors.

Nevertheless, since the suggested research seeks to improve the quality of the data and thus frequently should increase the amount of information reported, the motivation of the respondent is crucial to the success of the innovations. Most of the suggestions below refer primarily to the respondent; the central role of the interviewer in their implementation must, however, also be considered. Respondent motivation depends on interviewer motivation, and the views and convictions of the field staff must be taken into account at all stages of the development of new field procedures.

Underlying all the recommendations is the belief that the data collection operation should be made more "respondent-friendly" and that the whole tenor of the exercise should be that of a cooperative enterprise in which both interviewer and respondent share the responsibility for the success of the interview. The ideas presented draw heavily on the work of Cannell. All of them imply a willingness to treat the respondent as a partner in the information gathering process and attempt to create an atmosphere/environment in which the interviewer and the respondent work together to satisfy the needs of the survey.

There was broad agreement among the participants that explanatory motivating instructions to respondents should be incorporated in the survey instrument; that the impact be assessed by experimenting with the inclusion of such instructions in different versions of redesigned questionnaires.

The use of a signed commitment by the respondent (with a signed undertaking of confidentiality by the interviewer) was suggested as a means of emphasising the need for complete and accurate reporting and clarifying the objectives and requirements of the survey to the respondents.

Reminder telephone calls each day for the CE diary were proposed as a means of reinforcing the importance of timely completion of the diary; a space in which to sign and date each day's entry in the diary was also recommended as an additional incentive to regular completion.

Financial incentives were recommended by some participants; my personal inclination is not to pursue that path.

The recommendations above are designed to emphasise to the respondent the importance not just of being a respondent but of being a good respondent; there is therefore the possibility that in some cases this emphasis will result in unit or item nonresponse. The relative weight to be given to response and nonresponse errors in evaluating the results of the fieldwork needs to be established if rational decisions are to be taken by the survey designers.

III. CONCLUSION

The suggestions for research involve different degrees of interference in the current survey process, ranging from the more efficient use of currently available data to the complete redesign of the survey instruments.

First, existing data provides, or could easily be modified to provide, further information on data quality. The overlap of the time periods for which information is collected at successive interviews gives an opportunity to check on omissions and telescoping; a development of the idea is described in this

report. For the diary the identification of items filled-in by recall at the time of the interviewer's visit would provide an opportunity to compare these reports with those filled-in by respondents before the interviewer arrived. An initial analysis of proxy- and self-reports could be conducted using the currently available data. Differences between types of respondents and types of subjects would imply the desirability and possibly the direction of further investigation.

Second, some field experiments are needed to establish the effects of the present field strategy. The first set would compare the standard CES respondent rules with alternatives, taking into account costs, level of reporting and nonresponse. In parallel, research on the identification of "best" respondents would be desirable.

Third, experiments on cognitive aspects of the survey response process would help the researcher to frame questions in such a way that it is easier for the respondent to respond accurately. Among the methodologies suggested were think-aloud interviews; behavior-coding pretest interviews; observation of respondents; and retrospective protocols.

The end-result of the research to improve our understanding of the survey response process should be a number of insights into respondents' behavior which can be used to generate hypotheses regarding effective methods for eliciting the desired information. These hypotheses can be used, in turn, to generate a set of survey items and instructions (or, better yet and even more likely) multiple sets of items and instructions, that can be evaluated for their ability to improve data quality (assuming, of course, that criteria for evaluating the quality of survey data have been developed).

Fourth, the importance of motivating the respondent to cooperate was discussed; suggestions ranged from the inclusion of explanatory motivating instructions to the use of a signed commitment form.

There are four underlying issues which should be borne in mind in considering the recommendations for research made in this report.

- (i) Objective measures of quality: Without such measures it is impossible to reach any defensible conclusions about the efficacy of the present procedures or the impact of changes in those procedures. In addition to measures of data quality, the implications of changes for unit and item nonresponse must be considered.
- (ii) Flexibility: In a complex survey of this kind it is unreasonable to assume that the same procedures or the same criteria should be used for all sections of the questionnaire or for all kinds of respondents. Any attempt to treat the whole survey as an indivisible

seriously sub-optimal results.

- (iii) Field Implementation: The field staff, and particularly the interviewers, are the critical intermediaries between the survey design and the respondent. No procedure, no matter how good in theory, will be of use unless it can be translated into a practicable set of instructions for the interviewers. By the same token, the expertise and knowledge of the field staff provide an essential input for decisions on changes in and improvements to the survey instruments.
- (iv) Experimentation: The recommendations made are not necessarily suitable for immediate incorporation into the survey instruments. The ongoing work on the survey can however provide a useful control group for evaluating experimental procedures. Judicious choice of samples for the experiments would provide a relatively low cost opportunity for comparisons.