

RESEARCH REPORT SERIES
(*Survey Methodology* #2010-07)

**A "Cognitive" Interviewing Approach for the Survey
of Income and Program Participation: Development of
Procedures and Initial Test Results**

Jeffrey C. Moore
Karen Bogen
Kent H. Marquis

Statistical Research Division
U.S. Census Bureau
Washington, D.C. 20233

This paper was previously published in November of 1992, in the Proceedings of Statistics Canada Symposium 92: Design and Analysis of Longitudinal Surveys.

Report made available online: June 9, 2010

Disclaimer: This report is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

**A "Cognitive" Interviewing Approach for the Survey of Income and Program Participation:
Development of Procedures and Initial Test Results**

J.C. Moore, K. Bogen, and K.H. Marquis¹

ABSTRACT

This paper describes the development and initial testing of experimental data collection procedures for the Survey of Income and Program Participation (SIPP). The new procedures derive from prior research which has revealed serious levels of measurement error in some of SIPP's basic statistics, the important implications of the errors for standard analytical uses of the data, and which has suggested the cognitive bases of the errors. The key features of the redesigned procedures are a clear and consistent message to all participants that accuracy is the primary goal, and an emphasis on the use of records to assist income reporting. Initial results from small-scale tests of the new procedures indicate a high rate of record use to report income flows, and decreased response error (as indicated by a reduction in underreport errors and in the "seam bias"); on the negative side, the initial tests have suffered substantially greater nonresponse than does standard SIPP, and possibly increased per-case costs.

KEY WORDS: Cognitive research; Measurement error; Questionnaire design; Record use; Seam bias

1. INTRODUCTION

1.1 The Survey of Income and Program Participation

The Survey of Income and Program Participation (SIPP) is a major, continuing demographic survey program of the U.S. Census Bureau, and an important source of key social and economic indicators for the United States. This large-scale survey provides the most comprehensive information ever assembled on the economic situation of persons and families in the United States. SIPP data contribute to a wide range of policy decisions—health insurance and pension coverage, tax reform, Social Security costs, the effectiveness of state and federal assistance programs, etc.

In its current design, a new SIPP panel is introduced every year, and has a life of about 2½ years; households in each panel are interviewed eight times at four month intervals. All household members aged 15 and older are eligible for interview. Each interview (or "wave") gathers monthly data for the four calendar months preceding the interview month. Self-response is the preferred reporting mode, but proxies are accepted for persons not available at the time of the interviewer's visit. Until recently, all SIPP interviews were conducted by personal visit. Starting in February 1992, however, interviews 3, 4, 5, 7, and 8 were designated as telephone interviews.

1.2 Overview of the Paper

This paper presents an interim report on a research program still in progress. It describes the SIPP Cognitive Research (SIPP-CR) Project, the goal of which is to develop and test alternative measurement procedures for the SIPP to reduce important measurement errors. Section 2 presents a brief description of prior research leading up to the current study. Section 3 describes the major features of the new procedures, and the ways in which they differ from standard SIPP. Section 4 outlines the research plan for the SIPP-CR Project, and Section 5 summarizes results of initial pretests using

¹ Center for Survey Methods Research, U.S. Bureau of the Census, Washington, DC, USA, 20233-4700. The views expressed herein are the authors' and do not necessarily represent the official views of the Census Bureau.

the new procedures. Section 6 offers some brief conclusions and thoughts about next steps. Additional information on these topics is presented in Marquis, Moore, and Bogen (1991).

2. PRIOR RESEARCH ON SIPP MEASUREMENT ERRORS

2.1 The "Seam Bias"

Previous research has revealed important measurement error problems in SIPP. In an early study, Burkhead and Coder (1985) identified a "seam bias" in the measurement of month-to-month change using SIPP data. The seam bias is the tendency for many more changes (for example, from "on" to "off" participation in some transfer program) to appear between adjacent months at the "seam" between two interview waves, than between two adjacent months within the reference period of a single interview wave. No reasonable scenario for true month-to-month change could produce this sort of pattern; thus, the seam bias is a clear indicator of problems in SIPP's measurement of change.

2.2 The SIPP Record Check Study

The SIPP Record Check Study (Moore and Marquis (1989), Marquis and Moore (1990)) was implemented to investigate the nature and extent of response error in SIPP, and, specifically, to better understand the nature of the seam bias and its causes. The study used a full-design record check to assess measurement quality for reports of participation in and income received from eight government transfer programs, in four states, for the first two waves of the 1984 SIPP panel.

The Record Check Study showed that reporting errors in SIPP are quite rare; overall, fewer than 2% of the reports about program participation or changes in program participation were found to be in error. The study also showed, however, that even low levels of response error can have severe effects on important estimates, both univariate estimates and measures of association. For program participation rates, Marquis and Moore (1990) report net underestimates in the 10-40% range. Program participation change rates are underestimated by even greater amounts within a wave (off-seam), while change rates on the interview seam are severely overestimated.

While the record check research has permitted detailed descriptions of SIPP response errors, it has proved much less useful in identifying causes of the response errors. Marquis and Moore (1990) examined several of the traditional hypotheses about the causes of survey response errors--forgetting, memory decay, confusion, proxy bias, etc.--and found that none was strongly supported by the data.

2.3 Exploratory SIPP Cognitive Research

In a further search for causal insights, the Census Bureau implemented a small-scale, exploratory cognitive research project to look for clues to the survey's response error difficulties in respondents' understanding of SIPP tasks and questions, and their thought processes in answering those questions. Census Bureau headquarters staff received training in "cognitive interviewing" techniques, and accompanied experienced SIPP interviewers while they administered the standard SIPP interview. The observers were free to interject questions to find out how the respondent interpreted a general task or a specific question, formulated a response, etc., or simply to observe the interviewer-respondent interaction. This project and its results are summarized in Marquis (1990).

The exploratory cognitive research yielded several important insights into the response error dynamics of the SIPP interview. One key insight was the limited role that memory plays in respondents' supposed "recall" of their reference period income. As a substitute for detailed, direct recall of the 4-month payment history, respondents tend to rely on very simple rules, combined with a few recalled facts, to construct a plausible (though not necessarily accurate) story about their income. Furthermore, this shortcut, "story-telling" strategy appears to be not only tolerated by current SIPP procedures, but encouraged in subtle but important ways. For example, interviewers' performance evaluations are primarily based on their response rates and efficiency. This environment can work against response quality, because it can discourage interviewers from "pushing too hard" for accurate answers, either through record use or through difficult, more complex recall strategies; from following up "don't know's" and item refusals; or from responding to, or even recognizing, respondents' elaborations to too-simple answers. The evaluation environment can also encourage interviewers to rush through the interview, and even to actively assist respondents in finding easy approximations to complex truth.

The cognitive research also pointed out many ways in which the current SIPP questionnaire contributes to response quality problems. In many areas, the questionnaire presents demands on memory that are simply unreasonable (for example, asking for highly unrealistic detail, or for the recall of material that is unlikely to have been stored in memory in the first place); in other areas respondents are effectively denied an opportunity to report accurately in the interest of processing efficiency (for example, in the requirement to report all income in monthly "chunks," even income which comes on a schedule not immediately translatable to monthly). These shortcomings force respondents away from reporting accuracy, and toward "story-telling." The SIPP questionnaire also falls short by failing to provide clear and consistent information to respondents about the nature of their task. The cognitive interview observers often found that respondents did not understand the point of an entire question series. This was sometimes due to the lack of explanatory transition statements between major topic areas; in other instances the complexity of the instrument, with its myriad of screeners and check items (often read aloud by interviewers, further disrupting the flow and context), was clearly at fault. The instrument also fails to provide adequate or consistent information about the level of accuracy or effort expected of the respondents.

3. AN ALTERNATIVE MEASUREMENT DESIGN FOR SIPP

The exploratory cognitive research provided important insights into the likely causes of SIPP's response error problems, and led directly to many of the changes incorporated into a set of alternative measurement procedures for the survey. The major components of the new procedures are as follows:

- The cornerstone of the new measurement procedures is the emphasis on respondents' use of personal income records to assist income reporting. Accurate recall from memory of a four-month stream of income is usually very difficult, and often virtually impossible; the revised procedures explicitly recognize this fact. They attempt to take the reporting task out of respondents' heads entirely (and thus are not really "cognitive" at all) by insisting instead that respondents use their personal records to report their income, in order to preempt their use of overly simple response strategies, and to ensure accuracy. Interviewers are also responsible for training respondents how to interpret their records, and how to maintain them for the next interview. This includes giving respondents a file folder for storing records between interviews, and, for income not accompanied by any record, a sheet on which to record the relevant details concerning that income (date, amount, source, and recipient).
- In the absence of income records, interviewers are trained to recognize unacceptable shortcut strategies, and to guide respondents to use more realistic recall strategies. In such circumstances, respondents are first asked to describe the "usual" pattern of payment dates and amounts; then to list factors that can conceivably affect payment dates or amounts; next to consider whether any of the possible "change" factors occurred during the reference period, and if so, when; and finally, from this complex mix of information, to reconstruct what actually happened during the reference period.
- To avoid "story-telling," to reinforce the message that accuracy is the primary goal of the survey, and to make records easier to use, the new procedures collect individual, "to-the-penny" income payments, not monthly totals. Regardless of how often respondents receive income from a particular source, interviewers collect dates and amounts for each individual payment. Monthly totals are produced by computer, not in respondents' heads. Even income sources for which data users may not need exact amounts are collected with the same level of precision, in order to ensure a consistent message to respondents that accuracy is vital, and that estimates are not acceptable.
- The new procedures use unstandardized interviewing techniques in the collection of income information. The alternative SIPP interview begins with a "free recall" section which attempts to clearly set out the goals of the section, and then allows respondents substantial control over the reporting of their income for the reference period. There is a structure to this part of the interview--the information goals are explicit, and the data capture mechanism gives clear guidance about the specific data needed. What is missing is the inviolable script, with pre-set questions in a pre-set order. There are many potential benefits of this format. By allowing respondents to report salient facts about their income with little delay, in the most natural order, without having to endure long strings of inapplicable or seemingly irrelevant questions, it enables them to become immediately involved in the interview and in the production of good information. It also equips the interview with the flexibility to handle the great complexity and diversity of people's income situations. Moore, Bogen, and Marquis (1992) provide a thorough description of this aspect of the revised SIPP procedures.

- The new procedures attempt to simplify the reporting tasks as much as possible, and clearly explain to respondents the purposes and goals of each section. Items have been re-ordered to make sections of the questionnaire more coherent. This change, as well as the "free recall" procedures described above, has eliminated the need for many complex skip patterns, allowing interviewers to concentrate on their essential task, which is no longer question-reading, but problem-solving. In some instances, the revised questionnaire opts to ask some items of a slightly larger-than-necessary universe of respondents, in order to eliminate preceding screener questions. Another change is the addition of short transition statements between the major sections of the questionnaire, to provide respondents with a guide for what to expect next.
- For the first interview, the new procedures insist on self-response, preferably "family-style," in a non-distracting interview setting. These components of the revised interview are intended to both reinforce the message that the survey seeks the highest quality information, and provide an environment that is most conducive to achieving high quality. In subsequent interviews, if the household has records available, the self-response and group interview rules can be relaxed; their purpose initially is to make sure all household members understand the goals and importance of the survey, to allow them to help each other recall income sources and details, and also to provide implicit approval for household members to share income information, thus paving the way for collecting high-quality proxy information (with records, of course) in future interviews.
- To directly attack the seam bias, in particular the overreporting of change at the seam, the revised procedures use overlapping reference periods with reconciliation of discrepant information, a technique adapted from Murray, *et al.* (1991). Unlike standard SIPP, each wave's reference period extends to the date of the interview, rather than ending on the last day of the full month preceding the interview. Since the reference period for the next interview starts at the beginning of the month in which the preceding interview took place, for interviews after the first there is an overlap period covered by both the current and the previous interview. At the second and subsequent interviews, the interviewer first collects income information independent of the previous interview, then reviews the information with respondents in light of previous information. There are two stages to this review. First, the interviewer resolves any discrepancies in income sources, checking all income sources reported in one interview but not the other for possible omissions. Following this, interviewers review both waves' data for discrepant income information in the overlap period, and resolve all discrepancies with respondents.
- An essential underpinning of the new procedures is a set of revised interviewer evaluation criteria, which are intended to encourage interviewers to attend to quality-oriented performance. The revised procedures no longer place primary and almost exclusive emphasis on high response rates and high efficiency, but add many indicators of the extent to which their performance is consistent with the primary quality goals, and raise those indicators to first-level importance. The main form of feedback is through monitoring a sample of tape-recorded interviews (all interviews are supposed to be taped) in such areas as obtaining group interviews and self-response, persuading respondents to use records, reconstructing income details in the absence of records using complex recall strategies, providing feedback to respondents, recognizing and solving respondents' difficulties, etc.

4. THE RESEARCH PLAN

The Census Bureau has designed a research program, currently in progress in the field, for evaluating and refining the revised "cognitive" procedures. This program includes two small pretests, a full-scale measurement quality evaluation study (now in the field), and an implementation research panel to address operational issues.

4.1 Pretest 1

The first pretest was conducted in Milwaukee, WI, from August through November 1991. Wave 1 interviews were conducted in August and September, with a standard four-month reference period; Wave 2 interviews, with a shortened two-month reference period², were conducted in October and November in households that had completed an initial interview two months before. The sample was 130 randomly selected addresses. The purpose of the first pretest was to assess the feasibility of, and refine as necessary, the new field procedures and instruments.

4.2 Pretest 2

The second pretest employed the same general design as the first: two months of Wave 1 interviews in December 1991 and January 1992, with a four-month reference period, and two months of Wave 2 interviews in February and March 1992, with a two-month reference period. The sample for Pretest 2 consisted of 130 individuals (and their associated household members), who resided in Milwaukee, and who were identified on official record systems as having received one of five income types--Aid to Families with Dependent Children (AFDC), Food Stamps, Unemployment Insurance, Supplemental Security Income (SSI), or earnings from a specific Milwaukee-area employer. The purpose of the second pretest was to test procedures for sampling from and matching to administrative and employer records; to develop data entry, database management, and data analysis strategies and programs; and to further test and refine the revised procedures and instruments.

4.3 The Evaluation Study

The Evaluation Study is currently underway, also in Milwaukee. When complete, it will include two waves of interviewing, each with a full four-month reference period. Wave 1 interviews began in September 1992, and will continue through January 1993; Wave 2 interviews will be conducted in February through May 1993. As in the second pretest, sample cases consist of individuals (and their associated household members) drawn from the record systems of one of five income sources. The goal is to complete approximately 350 Wave 2 interviews under each of two randomly assigned treatments: standard SIPP measurement procedures and the redesigned procedures.

The purpose of the Evaluation Study is to provide a direct comparison of measurement quality across the two treatments, using administrative and employer records as the primary criteria for assessing quality. Program (and employment) participation and amounts as reported by the respondents will be compared to the "true" information in the records. In addition, cost component comparisons (travel time, interview time, edit time, etc.) will be made across the two treatments to evaluate the costs of the new procedures, and to identify the causes of any cost differences. Lastly, in addition to a simple comparison of nonresponse rates, the record data will permit some comparisons of the characteristics of nonrespondents across treatments, which may provide an indication of nonresponse bias differences between the two treatments.

4.4 Implementation Research

If the Evaluation Study yields evidence of substantial quality improvements with the new procedures, with reasonable costs and reasonable nonresponse, further research will be conducted to address the many operational issues that will inevitably remain (for example, generalizability to other sites, respondent cooperation over multiple waves, use of computer assisted personal or telephone interviewing, differential effects on subgroups, costs and response quality effects of the individual components of the new procedures, etc.). The exact design and goals of this implementation (or operational) research have yet to be specified.

5. PRETEST RESULTS

The main purpose of the first pretest, and an important purpose of the second pretest as well, was to field test the new procedures and instruments, and to identify and correct the most obvious problems. While none of the basic features of the new procedures proved infeasible in the field (and several were surprisingly successful), throughout the pretests many refinements were made to the procedures and instruments as a result of situations encountered in the field and feedback from the interviewers. The second pretest was very informative about sampling from the various record systems, as a test of these procedures for the Evaluation Study. One important finding was the frequency with which the household roster for the address supplied by the agency/employer failed to include the target sample person². To account for this attrition, more sample cases were selected for Wave 1 of the Evaluation Study. Another goal of the second pretest was to test data entry procedures. That exercise, too, was very informative, pointing to the need for some important modifications for the next research phase.

The remainder of this section summarizes pretest results in three areas: the successful implementation of the new quality-oriented field procedures, indicators of improved measurement quality with the new procedures, and areas in which there is clear need for improvement--nonresponse and costs.

5.1 Implementation of the Quality-Oriented Procedures

Tape Recording Interviewers were reasonably successful in tape recording pretest interviews, although there is certainly room for improvement. In each test, about 75 % of all completed interviews were tape recorded. It is worth noting that, according to interviewers' reports, only one or two of the taping failures was attributable to respondents⁴. In almost every case, the failure to tape was due to mechanical failure, operator error, or the interviewer's failure to make the request of the respondent (which often happened in refusal conversion cases). These results offer fairly convincing evidence that tape recording is not a major issue for respondents.

On the other hand, the interviewer performance evaluation and enhancement system as a whole, the reason for the tape recording, did not work very well in the pretests. Objective evidence is scant, but there appear to have been major problems in the process of converting monitoring results into effective interviewer performance feedback. One problem was turnaround time, which was often much too protracted. A more basic difficulty, and one for which operational solutions are not immediately apparent, is interviewers' negative reactions to the monitoring. Our intent was to provide a continuous, on-the-job training system that would assist interviewers in improving their performance; interviewers, however, tended to be blind to any positive side to monitoring. Some felt that the system failed to take account of all of the interview interactions not amenable to audio taping, and viewed it mainly as a way to tally and document their errors. Interviewers did acknowledge, however, that the monitoring forms conveyed very clearly the highest priority interviewing goals and the behaviors in which we were most interested.

Group Interviews and Self-Response The group interview and self-response procedures appear to have been quite successfully implemented in Pretest 1 (data are not yet available for Pretest 2). Three-fourths of all interviewed adults who lived in multiple-adult households participated in a group interview, and 92 % of all interviewed adults self-responded. Standard SIPP procedures typically yield about 65 % self-response. (Standard SIPP only allows individual interviewing, so there is no comparable group interview figure available for comparison.)

5.2 Indicators of Improved Quality

The most definitive evidence of data quality derives, of course, from the matching of survey data with the administrative and employer records. A limited set of such matched survey/record results is currently available from Pretest 2. These results, as well as two other sets of analyses drawn from both pretests--respondents' use of records, and a reduced seam bias--suggest that the revised procedures do yield improved data quality.

Record Use Respondents' use of records in the pretests far exceeded expectations. At the household level, 87 % percent of all households (in both pretests combined) produced at least one record to assist income reporting, with very little difference between Wave 1 and Wave 2. Record use at the income source level was 72 %--that is, for 72 % of the income sources respondents reported, at least one record was used to substantiate the date and amount of a payment. Similarly, at the payment level, respondents used records to report 63 % of their individual payments. The Wave 2 payment-level record use rate of 74 %, versus 57 % in Wave 1, again suggests that, although there is still substantial room for improvement, interviewers successfully trained respondents in record maintenance between interviews⁵.

Standard SIPP procedures also encourage interviewers to ask respondents to use records. According to results summarized by Singh (1991, 1992), the record use rate at the income source level was about 20 % in the initial waves of the 1991 SIPP panel. Regular SIPP's rather limited success in this regard may be in part attributable to interviewers' fears that asking for records will irritate respondents, causing breakoffs and subsequent nonresponse, and will also increase interview time, thus lowering their efficiency.

Seam Bias More direct evidence of improved quality with the revised SIPP procedures is apparent in an analysis of the seam bias. Table 1 shows an overall "seam bias index"--the ratio of the average number of month-to-month changes on the seam to the average number off the seam--for each pretest, collapsed across all income types. An index value of 1.0 indicates no seam bias; that is, the index is 1.0 if the number of transitions measured at the seam is the same as the number of transitions in an average off-seam pair of months. For the first pretest, the overall seam bias index is .95; for Pretest 2 it is a slightly higher 1.55, still substantially lower than the results reported for standard SIPP by Burkhead and Coder (1985)⁶.

We can speculate about why the new SIPP procedures appear to result in a more even distribution of reported change. Marquis and Moore (1989) have shown that the seam bias is the net result of both an underreporting of changes within an interview (off the seam) and an overreporting of changes across interviews (on the seam). The focus on individual

payments in the new procedures may encourage respondents to report income receipt in all (or at least more of) its messy detail. Regular SIPP procedures, because of their focus on monthly aggregates, push respondents away from details, and toward telling a plausible, summary story. At the next interview, respondents may tell a slightly different plausible story; by this process, change may be minimized within an interview, and forced to appear at the seam. Other procedural changes which may also have contributed to the seam bias reduction are the use of overlapping reference periods, the resolution of income source discrepancies across interviews, and the resolution of differences in reported income receipt during the overlap period. It must also be noted that the seam bias reduction is to an unknown extent an artifact of the design of the pretests, which used a shortened, two-month Wave 2 reference period instead of the four-month reference period of standard SIPP.

Underreport Errors As noted above, there is direct evidence of the measurement error effects of the SIPP-CR procedures in the administrative and employer record data available in Pretest 2. To date, the survey reports of known program participants have been evaluated against record-based "truth" for two programs, Food Stamps and Supplemental Security Income (SSI).

Table 2 presents monthly program participation underreporting error rates—that is, the proportion of "true yes" months of participation which respondents failed to report in the survey⁷. Because of the small sample size of Pretest 2 and the vast differences in design between the pretests and standard SIPP, we have not attempted any statistical tests, and thus make no claims concerning the statistical significance of the observed differences. Nevertheless, the limited evidence again suggests that the revised procedures are moving in the right direction with regard to making important improvements in the measurement of key SIPP statistics.

5.3 Areas Needing Improvement - Household Nonresponse and Costs

The pretests were not designed to provide definitive operational comparisons to standard SIPP procedures. However, pretest data suggest that, as currently designed and implemented, the new procedures may fall well short of standard SIPP performance in two key areas—nonresponse and costs.

Nonresponse Across both pretests combined, the Wave 1 household response rate (the number of interviewed households divided by the number of eligible households) was 73%; the rate for Wave 2 (based only on Wave 1 interviewed households) was 87%, yielding a longitudinal response rate of 63%. This rate indicates the proportion of Wave 1 eligible households that were interviewed in both waves. Regular SIPP achieves a Wave 1 response rate of about 92%, and a longitudinal rate at Wave 2 of about 88%. While SIPP's rates are not exactly comparable to the pretests' due to procedural differences (for example, regular SIPP follows movers, the SIPP-CR pretest procedures did not), it is quite clear that the pretest response rates were much lower from the outset at Wave 1, and that attrition is also likely to have been higher in Wave 2.

SEAM BIAS INDEX:

SIPP-CR Pretest 1:	0.95
SIPP-CR Pretest 2:	1.55

Representative Seam Bias Indices for Standard SIPP (Burkhead and Coder, 1985):

Unemployment Comp:	1.9
Earnings:	2.2
Food Stamps:	3.5
Social Security:	3.9
AFDC:	4.9
Private Pensions:	6.3

TABLE 1: Seam Bias Results for the SIPP-CR Pretests and Standard SIPP

MONTHLY PROGRAM PARTICIPATION UNDERREPORTING:

% of "true yes" months reported as "no":

	<u>SIPP-CR</u>	<u>Standard SIPP</u>
Food Stamps	9.7%	23.7%
SSI	11.1%	23.2%

(Standard SIPP results from Marquis and Moore, 1990)

TABLE 2: Program Participation Underreporting for the SIPP-CR Pretests and Standard SIPP

We reviewed interviewers' descriptions of the circumstances of each noninterview they encountered for evidence that the new procedures caused the higher nonresponse. With one or two possible exceptions, there is scant evidence in these reports that any noninterview was a direct result of the new procedures. Not-at-home noninterviews, which comprised about 20-25% of the noninterview cases, are unlikely to be due to any special survey procedures, certainly not in a first interview wave. The majority of noninterviews were refusals. In almost every case, Wave 1 refusals happened before the interviewer could even begin to explain the purpose of the survey and what was involved. Although Wave 2 refusals by definition occurred with knowledge of what the interview held in store, even those refusers, according to interviewers' reports, did not implicate any of the cognitive procedures in their refusal behavior. People did not refuse because they were asked to get records or because they were going to be tape recorded. The pretest nonresponse problems appear to have been much more administrative in nature; potential refusals and difficult-to-locate respondents were often not identified soon enough to take effective corrective action, or, if they were identified early, followup action was often not immediate.

Costs While it is difficult to compare the SIPP-CR pretest costs directly to the costs for regular SIPP (due to much smaller assignments in SIPP-CR, for example, and a highly clustered sample design for regular SIPP), it is quite clear that the SIPP-CR pretests experienced substantially higher field costs than those associated with the standard administration of the survey, perhaps as much as 50% higher. An obvious hypothesis is that some of the features of the new procedures--maximum self-response, group interviews, insistence upon an appropriate interview setting, the use of records, etc.--were responsible for the cost increases, since they required many additional visits to the households that would have been avoided under standard SIPP procedures.

We reviewed interviewers' reports of their visits to Pretest 1 Wave 1 households, all of which were supposed to have been recorded, and subjectively judged whether each would have been necessary under standard SIPP procedures, or whether it was an "extra" contact, required only to carry out the new procedures. All first visits, for example, were classified as non-extra's; all visits to obtain missing income records were "extra." The Pretest 1 visit record data do not show an unreasonable number of "extra" household visits (data from Pretest 2 have yet to be analyzed). Although an exact count of the number of extra visits is impossible, an upper limit can be assessed; we estimate that at most 14% of all Wave 1 personal visits to interviewed households were extra. While these extra visits (and the many extra telephone calls that would not have been necessary under standard SIPP procedures) undoubtedly contributed to higher field costs, they do not seem sufficient in number to explain the full cost differential.

Another contributor to the higher costs of the new procedures is actual in-house interviewing time. For the second SIPP-CR pretest, a Wave 1 interview took an average of 71 minutes per household; for regular SIPP, the average is about 52 minutes per household. This difference may be attributable to interviewer inexperience (all SIPP-CR interviewers were new to the job), or it may be due to the procedures; in either case it is unlikely to have contributed greatly to the observed cost difference.

A much clearer major cause of the higher pretest field costs was the fact that the interviewers made many unproductive visits (Krasko, 1992). There was a clear avoidance of interviewing in the evenings, so interviewers made repeated daytime visits that did not yield any contact with potential respondents. Since travel costs are a major component of field costs, these non-productive visits undoubtedly contributed to the higher direct interviewing costs. It is possible that the interviewers' inexperience as survey interviewers, the fact that they did not live in their assignment areas, and the lack of emphasis on costs and efficiency (in training, supervision, and feedback), all contributed to the interviewers' making so many non-productive visits.

6. CONCLUSIONS AND NEXT STEPS

Although work on the revised, "cognitive" SIPP procedures is still very much in progress, indications from small-scale pretests are that the new procedures have the potential to substantially reduce some of the survey's important measurement problems. At the same time, the operational difficulties encountered in the pretests--high nonresponse and high costs--clearly put at risk the notion that they are a viable option for national, production implementation.

The Evaluation Study currently underway--a side-by-side experimental comparison of the new procedures and standard SIPP procedures, using administrative record data as criterion measures--will yield solid evidence about the measurement error benefits of the revised procedures. Should the measurement error results prove sufficiently positive, the Census Bureau will conduct additional research to address the many operational issues that will remain, including, of course, how to bring nonresponse and costs under control, but also the generalizability of the results to other sites, respondent cooperation over multiple waves, how best to exploit computer assisted personal or telephone

interviewing with the new procedures, the differential effects of the new approach to gathering income data on subgroups (especially high income subgroups), and many other issues.

Among these "other" issues, two deserve special mention. One has to do with interviewer behavior, which the new procedures clearly push in new directions. A reasonable interpretation of the pretest nonresponse results is that while respondents show little reluctance about cooperating with the new procedures, interviewers may well be signaling some reluctance to administer them. Our evaluation of the results of the Evaluation Study must be attentive to interviewers' perceptions: what, if anything, do they find particularly onerous about the new procedures, and why? We must also be prepared to accept the possibility that the new procedures may require further revision and refinement to make them truly doable, as well as new classroom training methods, and new approaches to nurturing and supervising interviewers in the field.

The second concerns the package of SIPP-CR procedures themselves. The creation of the current cluster of procedures was driven by SIPP's redesign schedule deadlines. We were not allowed the luxury of time to develop and refine the components individually, but instead had to take an unquestionably "kitchen sink" approach. Although we can speculate, we do not know where in the package the real quality gains occur, and where the gains are not sufficient to justify the added expense. If increased costs and nonresponse continue to accompany any gains in response quality, it will be essential that the next phase of the research address the discrete cost and quality effects of the individual components of what is now the SIPP-CR measurement package.

NOTES

2. In both pretests, the Wave 2 reference period was shortened in order to allow the research program to meet survey redesign schedule deadlines. This aspect of the design of the pretests may have affected key results, especially those having to do with the apparent reduction in the seam bias (see Section 5.2).
3. Observed match rates—the rate at which the target sample person was found in the roster of household members in Wave 1 interviewed households—ranged from a low of 68% for AFDC to 96% for the employer and Unemployment Compensation samples.
4. The fact that Wave 2 nonresponse was higher in the SIPP-CR pretests than standard SIPP typically experiences (see Section 5.3) could suggest a negative response by Wave 1 respondents to the new procedures, including, possibly, the taping requirement. However, there is no explicit mention of a problem with tape recording in any of the Wave 2 noninterview reports.
5. A simple t-test under the assumption of sample independence is significant. Taking the correlation of the Wave 1 and Wave 2 observations into account does not change the conclusion drawn from the original test. Since some persons appear in only one wave, we re-estimated the record use proportions only including people who were in both waves. The results are very similar, so we conclude that using all available cases does not distort the difference conclusion importantly.
6. The SIPP-CR seam bias results are based on data from all households which completed both interview waves; 74 in Pretest 1 and 79 in Pretest 2. The data reported by Burkhead and Coder are from the first three interview waves of the 1984 SIPP Panel, comprising approximately 20,000 households.
7. SIPP-CR Pretest 2 sample persons experienced 165 true months of Food Stamps participation, according to the administrative records, of which they reported 149 in the SIPP-CR interview; for SSI, the comparable numbers are 135 true participation months, of which 120 were reported. The data reported by Marquis and Moore are from the SIPP Record Check Study, which used a three-state subset of the first two interview waves of the 1984 SIPP Panel. Eligible SIPP sample persons during this time period experienced 1,451 true months of Food Stamps participation, according to administrative records, of which they reported 1,107 in the standard SIPP interview; for SSI, the comparable numbers are 919 true participation months, of which 706 were reported.

REFERENCES

- Burkhead, D. and J. Coder (1985). "Gross Changes in Income Reciprocity from the Survey of Income and Program Participation." Proceedings of the Social Statistics Section, American Statistical Association, Washington, D.C., pp. 351-356.
- Krasko, N. (1992). "SIPP-CR Pretest I Type A Analysis." Unpublished U.S. Bureau of the Census memorandum for Stephen Willette, February 1992.
- Marquis, K. (1990). "Report of the SIPP Cognitive Interviewing Project." Unpublished report, U.S. Bureau of the Census, August 1990.
- Marquis, K. and J. Moore (1989). "Some Response Errors in SIPP--With Thoughts About Their Effects and Remedies." Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 381-386.
- Marquis, K. and J. Moore (1990). "Measurement Errors in SIPP Program Reports." Proceedings of the 1990 Annual Research Conference, U.S. Bureau of the Census, Washington, D.C., pp. 721-745. Also available as Report No. 9008 in the Census Bureau's SIPP Working Paper Series (June 1990).
- Marquis, K., J. Moore, and K. Bogen (1991). "A Cognitive Approach to Redesigning Measurement in the Survey of Income and Program Participation." Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 413-418.
- Moore, J., K. Bogen, and K. Marquis (1992). "Use of Unstandardized Interviewing Techniques in a Proposed Redesign of the Survey of Income and Program Participation." Unpublished paper prepared for the Joint Meetings of the Census Advisory Committees of the American Marketing Association and the American Statistical Association, U.S. Bureau of the Census, Washington, D.C., October 22-23, 1992.
- Moore, J. and K. Marquis (1989). "Using Administrative Record Data to Evaluate the Quality of Survey Estimates." Survey Methodology, Vol. 15, pp. 129-143.
- Murray, T. S., S. Michaud, M. Egan, and G. Lemaitre (1991). "Invisible Seams? The Experience with the Canadian Labour Market Activity Survey." Proceedings of the Annual Research Conference, U.S. Bureau of the Census, Washington, D.C., pp. 715-758.
- Singh, R. (1991). "SIPP 91: Wave 1 Results of the Record Check Study." Unpublished U.S. Bureau of the Census memorandum for the SIPP Research and Evaluation Steering Committee, December 19, 1991.
- Singh, R. (1992). "SIPP 91: Wave 2 Results of the Record Check Study." Unpublished U.S. Bureau of the Census memorandum for the SIPP Research and Evaluation Steering Committee, June 15, 1992.