THE SURVEY OF INCOME AND
PROGRAM PARTICIPATION

PRETEST RESULTS OF AN ALTERNATIVE MEASUREMENT DESIGN FOR THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

No. 165

K. Bogen, J. C. Moore and K. H. Marquis
Center for Survey Methods Research and
Census Bureau

U. S. Department of Commerce BUREAU OF THE CENSUS

ABSTRACT

This paper describes the development, design, and initial testing of experimental data collection procedures for the Survey of Income and Program Participation (SIPP). The new procedures derive from prior research indicating the amount of measurement error in some of SIPP's basic statistics (and the important implications of the errors for standard analytical uses of the data), and suggesting the cognitive bases of reporting errors. The key features of the redesigned procedures are: a clear, consistent message to all survey participants that the primary goal is response accuracy; and a recognition that because accurate data are not easily accessible in memory, one needs to use income records. Initial results from a small-scale test of the new procedures indicate a high rate of personal record use to report income flows, and reduced response error (as indicated by a reduction in the "seam bias"); on the negative side, the initial test suffered substantially greater nonresponse than does standard SIPP, and possibly increased per-case cost.

ACKNOWLEDGEMENTS

The authors thank Nola Krasko of the Census Bureau's Field Division and Doug Lampe of the Census Bureau's Demographic Surveys Division for their important contributions to this research. In addition, we acknowledge the work of the other members of the Census Bureau's SIPP Cognitive Research Committee: Don Fischer, Bob Kominski, Nancy Mathiowetz, Tom Palumbo, Gary Shapiro, Kathy Short, Lynn Weidman, and Steve Willette.

I. INTRODUCTION

The Survey of Income and Program Participation (SIPP) is one of the Census Bureau's major, continuing demographic surveys. It was initiated by the Census Bureau in 1983 and has emerged as one of the Nation's key social and economic indicators. 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, and many others.

This paper describes the SIPP Cognitive Research (SIPP-CR) Project, a Census Bureau research project to develop and test alternative procedures and field materials for the SIPP, in an effort to reduce important measurement errors. This paper covers the background information and previous research that led up to the current research, the differences between the test procedures and regular SIPP, results of the first pretest of the experimental procedures, and the plans for the remainder of this research project. (Additional information on these topics is presented in Marquis, Moore, and Bogen (1991).)

II. BACKGROUND

Previous research has indicated that there are important errors in the SIPP data. In some early research on SIPP, Burkhead and Coder (1985) identified a seam bias. They found that respondents' monthly reports show many more changes on or off government programs, as well as changes in amounts received from those programs, in two adjacent months between reference periods (on seam) than between two adjacent months within the same reference period (off the seam). Figure 1 illustrates the definition of the seam. Because of the staggered "rotation group" interview design, all month-pairs throughout the calendar year are equally represented among all interview seams; consequently, there should be the same number of changes between months, regardless of whether the months are within or across reference periods.

The SIPP Record Check Study (Moore and Marquis (1989), Marquis and Moore (1990)), used a full record check design in which respondents were administered the regular SIPP questionnaire. Response information provided about participation in eight government programs (four Federally-administered and four State-administered programs) was checked against the administrative records for that program in four states.

One major finding of the Record Check Study is that errors are rare overall. Less than 2% of the information about program participation or changes in program participation was found to be wrong. However, since true participation and true participation changes are also rare, these few response errors cause important biases in estimates made by SIPP data users. Levels of participation were found to be underestimated by 10-40% for some programs. Change rates off the seam were underestimated by even greater amounts, while change rates on the interview seam were severely overestimated. These errors, though rare, were found to have possibly severe effects on relationship estimates, which could make it extremely difficult to detect even strong associations.

The record check data were also reviewed in an effort to identify the causes of the errors in SIPP. However, Marquis and Moore (1990) found that none of the traditional hypotheses about the causes of survey response errors--forgetting, memory decay, confusion, proxy bias, etc.--were strongly supported by the data. In the end, the response errors found in the SIPP data could not be explained by any of the standard hypotheses about the causes of response errors.

As a result, the Census Bureau implemented a small-scale, exploratory cognitive research project to try to gain a better understanding of respondents' thought processes in answering SIPP questions. Several Census Bureau headquarters employees were trained in techniques for eliciting thought processes during interviews and accompanied SIPP interviewers to nonsample households where respondents were administered the standard SIPP interview. The observers were free to interject questions to find out how the respondent interpreted a question, formulated a response, or anything else. The researchers tried to learn respondents' understanding of the task, what, if anything, the respondent actually recalled, and how the respondent decided to answer. This project and its results are summarized in Marquis (1990).

The cognitive interviews revealed that respondents tend to adopt very simple rules to quickly answer questions about a specific 4-month stream of income, and they use these rules as a substitute for detailed, direct recall of the 4-month payment history. Typically, respondents took a few recalled facts (such as that they took last week off), and combined those few facts with some very simple rules (such as that they usually work 40 hours a week), and from this they constructed a plausible (though not necessarily accurate) 4-month payment stream.

Furthermore, the researchers felt that the interviewers, as a result of their evaluation system, may actually cause or encourage the "shortcut" answering respondents use. Interviewers are rewarded for high response rates and efficiency, goals that may be inconsistent with the goal of obtaining the highest quality data.

Another finding of this research was that even though respondents fairly frequently misunderstood questions, their misunderstanding was not limited to a few questions. The observers felt that respondents often did not understand the point of an entire question series. For example, they may not have understood that the questions had moved off of the government programs topic and onto private health insurance. So, the questions did not make sense to the respondent because he or she had lost track of the context. Observers identified two likely causes of this confusion: first, the SIPP questionnaire does not have section transitions that come between large breaks in topics and that explain the goals and purpose of the upcoming section; and second, interviewers too often read the check items and interviewer instructions that precede new sections or come in the middle of a question series. Check items are supposed to be filled without asking, since the information is already known. When interviewers read a check item, it breaks the flow and context of the succeeding questions.

The researchers also identified other potential causes of response errors: the instrument fails to provide adequate or consistent information about the level of accuracy or effort expected of the respondents, the questions force respondents to recall income information in ways that are not consistent with

how they typically process and store income information, and the complex skip patterns often lead to high item nonresponse rates.

The preliminary cognitive research provided important insights into the causes of SIPP response errors and led directly to many of the procedural changes that were made for the alternative design, described in the next section.

III. THE ALTERNATIVE MEASUREMENT DESIGN

The alternative measurement design for SIPP includes new field materials and new procedures. All of the key new procedures were introduced as a direct result of the findings from the SIPP Record Check Study and the cognitive research. Following are the major components of the new procedures, including brief descriptions of how they contrast with regular SIPP procedures, and the intent of the change:

A. Use Personal Income Records - In an effort to preempt respondents' use of simple strategies for "recalling" their 4-month income history, the revised procedures insist that respondents use their personal records to report their income. The idea of using records is introduced as the usual procedure, the norm not the exception.

Under the new procedures, the interviewers also provide respondents training about keeping and interpreting records. This includes providing respondents with a customary place to store their records between interviews, in this case a folder with SIPP information and literature, and a sheet on which to record income information for sources that do not come with any sort of record.

B. Collect "To the Penny" Income Information - The interviewers are given clear instructions, which they are to convey to respondents, that the primary goal of the new procedures is complete and precise, "to the penny" income information. The instrument is designed so that the accuracy message is consistent throughout the interview. Even income sources for which data users are not as interested in exact amounts are collected with the same concern for precision, in order to make sure that the message that estimates are not acceptable is given consistently to respondents.

C. Collect Individual Income Payments - Another major contrast with the standard SIPP procedures is the units in which the amounts information is collected. Regular SIPP collects monthly amount totals for most income sources, and asks for some asset amounts over the entire 4-month reference period. In order for respondents to report most accurately using their records, the experimental procedures ask for individual income payments. If a respondent is paid every week, the interviewer collects the amount for each payment in the reference period. Likewise, if the person is paid monthly, or interest on an account is received monthly, then that is how the interviewer collects and records the information.

An additional purpose of asking for payment-by-payment information is that it helps to send the consistent and clear message that accuracy is very important, and that respondents are not to report a "typical" story

but to report all the disorderly facts. When regular SIPP collects summary amounts, this may inadvertently send the message to respondents that estimated or plausible amounts are good enough.

D. Use More Realistic Recall Models - When respondents do not have records, which the new procedures anticipate may often be the case in initial interviews, interviewers are to supplement the respondents' use of shortcut methods for reporting their 4-month income streams. Interviewers are trained to guide respondents to use more realistic models to recall the dates and the amounts of their income payments during the reference period for that income source. Respondents are asked to consider what determines the size of each gross payment, and what would make the amounts of the payments vary, such as hours worked, leave taken, overtime pay, or a varying number of days in the payment period. Likewise, they are asked what would make the payment date vary, such as a usual payday falling on a weekend or holiday. Interviewers are to be alert to strings of payments that are all identical, and to probe for details of any variation from the typical situation.

E. Use an Unstandardized Approach to Collect Income Data - The instrument for the alternative measurement design is somewhat different than the standard SIPP. The alternative interview begins with just the basic demographic questions to ascertain the minimum necessary information to start the interview, mostly household roster information. Then the interview moves directly on to its main task--collecting complete and accurate income information. Respondents have been told that this is the goal, so the interview is arranged to get to it as quickly as possible. The remainder of the usual battery of demographic questions is saved for a later part of the interview.

After the household roster questions, the interview starts with a "free recall" section in which respondents are reminded of the goal of the section, and are asked to get their records and start reporting their income for the reference period. This open-ended format allows respondents to report their income in the order in which it comes to mind. They are not constrained to provide their income sources in the order of the questions. (Figure 2 is the Free Recall item from the questionnaire, and Figure 3 is an updated version of the main worksheet that is used to record income information.) This free recall section also forces respondents to become involved in the interview. They cannot simply listen and nod and respond in one-word answers, but must be active participants in the interview.

The free recall section of the interview is followed by a "recognition" section in which the interviewer reads a list of the major income sources and asks the respondent if he or she had any of those income sources. The purpose of this section is to make sure that no income sources have been missed in the free recall section.

F. Use Overlapping Reference Periods with Reconciliation - In an effort to reduce the overreporting of change at the seam, the alternative procedures use an overlapping reference period, with reconciliation of sources and amounts reported in the overlap period. Each wave's interview reference period extends to the date of the interview; it does not end on the last day of the full month preceding the interview, which is

how regular SIPP defines the reference period. The reference period for the next interview for the alternative procedures starts at the beginning of the month in which the preceding interview took place and ends on the date of that interview. Thus, for the second and subsequent interviews there is an overlap period of varying length depending on how far into the interview month the previous interview was conducted.

At the second interview, the interviewer administers the free recall and recognition sections just as they are administered in the first interview. When those sections are completed, the interviewer resolves any discrepancies in income sources. That is, he/she asks respondents about all income sources that were reported in one of the interviews but not in the other, checking for possible omissions in either interview. Following this income source resolution, interviewers resolve any discrepant dates or amounts in the overlap period. All dates and amounts in that period should be identical; interviewers work with respondents to resolve all discrepancies.

G. Simplify and Clarify the Questionnaire - Many of the complex skip patterns that are in the regular SIPP questionnaire have been removed so that the interviewer can concentrate on listening to respondents and helping them understand the items. The simplification of skip patterns was accomplished by using free recall of income (described earlier) which removes the need for many skips, by changing the order of some items and some sections of the questionnaire, as well as by asking some items of a slightly larger universe of respondents in order to eliminate a preceding screener question.

Another change to the instrument is the addition of short transition statements between the major sections of the questionnaire. In this way, respondents can get some idea of what to expect in the upcoming section.

H. Use a Non-Distracting Interview Setting and a Group Interview - To reinforce the message that the survey seeks high quality information, the new procedures call for an appropriate, non-distracting interview setting. It is not acceptable to conduct a "doorstep" interview with the new procedures. The interview has to be conducted in a setting that allows respondents to think and to have access to their records. The new procedures also call for a group interview format where possible, in which all eligible household members are interviewed together. The purpose of this feature of the design is to have household members help each other recall income sources, details about work schedules, and other income-related information. The group interview also demonstrates that household members are willing to talk about their income in front of each other, and gives at least implicit permission to collect proxy information in future interviews.

I. Revise Interviewer Evaluation Criteria - Although the system was not as effectively implemented in the pilot studies as had been planned, the evaluation of interviewers is another major contrast to standard SIPP procedures. The alternative procedures, as opposed to regular SIPP, put much less emphasis on high response rates and high efficiency. Interviewers are provided feedback on the extent to which their performance is consistent with the primary goal of obtaining complete and accurate

information from respondents. The main form of feedback is through monitoring of tape-recorded interviews. All interviews are supposed to be taped; the tapes are monitored and interviewers are evaluated on their efforts in the following areas: obtaining group interviews and self-response, getting respondents to use their records, reconstructing missing amounts using proper rules and not shortcuts, providing feedback to respondents on how they are doing, using transitions properly, and signaling to the respondent that he/she (the interviewer) is listening by not asking questions for which he/she already has the answer.

In summary, the experimental procedures attempt to convey to all participants that the survey is to be a collaborative effort between the interviewer and the respondents, who work together to report and record complete and accurate information about household income. All procedures, as well as the instruments, are designed to send a consistent and clear message that obtaining the highest quality data is the primary goal. Furthermore, the procedures and instruments are designed to aid the interviewer in achieving that goal.

IV. THE RESEARCH PLAN

The overall plan for the SIPP Cognitive Research (SIPP-CR) Project is to conduct two pretests of about 100 interviews each, followed by a full-scale evaluation study of about 700 interviews. All interviewing is to be done in Milwaukee County, Wisconsin.

The first pretest was conducted from August 1991 through November 1991. August and September were Wave 1 interviews (n=92), each using a standard four-month reference period, and October and November were second wave interviews (n=74), conducted in households that had completed an initial interview two months before. (The work was spread out over two months in order to provide the interviewers with a more constant, yet manageable, workload. The second interview was limited to a two-month reference period, instead of the usual four months, because of time limitations.)

The purpose of the first pretest was to assess the feasibility of, and refine as necessary, the new field procedures and instruments. For example, it was important to determine respondents' acceptance of a modestly unstructured interview as was planned with the free recall section. Likewise, it was equally important to determine if the information reported in an unstructured interview could be recorded in a usable manner. Among the other things to be tested in the first pretest were respondents' willingness to be tape recorded, their acceptance of the recognition section (after they had already reported their income in the free recall section), respondents' ability and willingness to find and use their personal income records, the logistics of arranging group interviews, as well as all the other procedural and instrument changes that were being tried for the first time.

The second pretest had the same general four-month design as the first pretest: two months of Wave 1 interviews (n=88) and two months of Wave 2 interviews (n=79). The second pretest was conducted between December 1991 and March 1992. The purpose of the second pretest was to test procedures for sampling from and matching to administrative record files, data entry, database management, and data analysis programs, as well as to further test the procedures and instruments.

The Evaluation Study is scheduled to begin in September 1992. The purpose and design of that part of the research program are described in Section VI of this paper.

V. RESULTS OF THE PRETESTS

The pretests were very successful in terms of achieving their primary goals. The main purpose of the first pretest, and a secondary purpose of the second pretest, was to test the new field procedures and instruments, and to identify and correct any problems with them. While none of the basic features of the new procedures proved to be 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 and cases that were found in the field and feedback from the interviewers. Likewise, the second pretest was very informative about sampling from administrative files, as a test of these procedures for the Evaluation Study. One important finding was that, for certain programs, many program participants cannot be found at the addresses provided by the administrative files[1]. Consequently, more addresses will need to be selected for the Evaluation Study. Another goal of the second pretest was to test data entry procedures. That, too, was very informative, pointing to the need for some important modifications for the next research phase.

Although work still continues on matching to administrative records, developing a database management system, and testing analysis programs, the second pretest has already provided invaluable experience in these areas as well. At a minimum, we will know what paths not to take; we fully expect, however, that out of the second pretest we will develop an integrated system that will ensure efficient analysis of Evaluation Study results.

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.

A. Implementation of the Quality-Oriented Procedures

One of the goals of the first interview was to get people to respond for themselves, since they were most likely to know how to obtain and interpret their records. An additional goal was to interview all eligible household members together, so that they could help one another recall their income sources as well as some of the details about receipt of income from those sources. While it is possible that the presence of other persons could deter some people from reporting certain sources of income, it was felt that the potential benefits of improved reporting sufficiently outweighed the potential losses to make group interviews the preferred format in the initial interview.

Group interview and self-response rates were quite good. Table 1 shows the pretest one results for self-response and group interviews: 75% of all

---

[1]Following are the match rates for the different sources. (The numbers are based on only the cases we are sure matched or did not match, and exclude any cases where match status was not known for sure.) AFDC=68% (n=22), Employer=96% (n=25), Food Stamps=83% (n=23), SSI=71% (n=24), and Unemployment Compensation=96% (n=24). The total match rate was 83% (n=118).

interviewed adults who lived in multiple-adult households participated in a group interview and 92% of all interviewed adults self-responded. Current SIPP gets about 65% self-response. (There is no group interviewing rule in SIPP.) The major reason for the difference in self-response rates between regular SIPP and SIPP-CR is in the callback rules. Regular SIPP allows (and even encourages) the interviewer to accept a proxy at the initial visit to the household; if a person is not home at the time of the interview, a proxy is taken. The SIPP-CR rules allow proxy reports in only a few situations (e.g., physical or mental incapacity, absence until after the end of the interview period, after two weeks into the month-long field period). The new procedures require interviewers to make a number of callbacks to get a self-response interview.

Another indicator that interviewers followed the quality-oriented procedures is in the high taping rate they achieved. Interviewers were to tape record all interviews. Table 2 shows that they taped about 75% of the interviews. According to field reports, virtually all of the interviews that were not taped were a result of operator or mechanical error or interviewer resistance, not respondent resistance.

B. Indicators of Improved Quality

While definitive evidence must await the matching of survey data with the administrative data, two sets of analyses--respondents' use of records, and a reduced seam bias--suggest that the revised procedures and instruments result in improved data quality.

Table 3 indicates the extent to which respondents used their personal income records to report the dates and amounts of the income they received. The record use rate at the household level was extremely high. Eighty-seven percent of all households used at least one record. There was very little difference between Wave 1 and Wave 2 for household record use. Record use rate at the income source level was 72%. This means that for 72% of the income sources, at least one record was used to substantiate the date and amount of a payment.

The results in Table 3 suggest that interviewers enjoyed some success at teaching respondents to keep their income records for the next interview. The observed source-level record use rate in Wave 2 was 78%, versus 67% in Wave 1.

Similarly, at the payment level, respondents used records to report 63% of their individual payments. The Wave 2 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 interviews.[2]

---

[2]Since reviewers have asked whether the differences are statistically significant, we did 2 analyses. First, we did a t-test as if the means were from independent samples. The difference is significant. However, there are two problems: 1) the Wave 1 and Wave 2 observations are correlated for people in both waves. Taking into account this correlation would not change the conclusion we drew from the original t-test; 2) some people are not in both waves. We re-estimated the proportions for only the people in both waves. The results were very similar, so estimating differences using all available cases does not appear to distort the change importantly.

Standard SIPP also encourages interviewers to ask respondents to use records. However, not all of the pretest record use results can be compared to regular SIPP because there are differences in how the income data are collected and the way the record use data are collected. Regular SIPP does calculate a record use rate at the income source level, which was about 20% in Wave 1 and in Wave 2 for the 1991 panel (Singh, 1991 and Singh, 1992), compared to 67% and 78% for Pretest Waves 1 and 2, respectively. Regular SIPP's limited success with respondent records 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. SIPP-CR does not place as much emphasis on those rating factors, which allows the interviewer to encourage record use without perceiving that he/she may jeopardize his/her own ratings; in fact, the SIPP-CR procedures are intended to explicitly encourage and reward record use.

Another strong indicator of improved quality with the revised SIPP procedures was a decrease in the seam bias. Table 4 shows the pretest one data indicating reduced bias in the reporting of participation status changes. The analysis uses a "seam bias index" - a ratio of the average number of month-to-month changes on the seam to the average number off the seam. It is an index to equalize the fact that there are more off-seam pairs of months than on-seam pairs of months. An index value of 1.0 means an equivalent number of transitions were measured in on-seam and off-seam pairs of months, or no seam bias.

For the first pretest, the overall seam bias index is .95, which means that, essentially, there were the same number of transitions reported from month to month regardless of whether the adjacent months were on or off the seam. (The data from the first pretest were collapsed across all income types since there were not enough cases to look at individual income sources.) Burkhead and Coder (1985) show status changes for the regular SIPP for selected individual income sources; Table 4 also summarizes the Burkhead and Coder results. Burkhead and Coder data result in indexes ranging from 1.9 for Unemployment Compensation to 6.3 for private pensions. In all cases, they find many more program status changes on the seam than off the seam. While their ratios are not directly comparable to the overall ratio of .95 from the first pretest, it is clear that regular SIPP procedures yielded a much bigger difference in reporting changes on the seam than off the seam.

We can speculate about why the new procedures result in a more even distribution of reported changes. Marquis and Moore (1989) have shown that the seam bias is a result of both underreporting of changes within an interview (off the seam) and an overreporting of changes across interviews (on the seam). The use of payment-by-payment information was implemented in the new procedures to focus respondents on actual receipt of income, rather than on a summary report. Regular SIPP forces respondents to tell a plausible, summary story when they provide information in one interview. When these people are interviewed again, they may provide a slightly different plausible story, again because the regular SIPP interview only collects monthly summary amounts and does not push for the details contained in income records. The changes then appear at the seam because respondents change their plausible story a little from one interview to the next. The SIPP-CR focus on the payment-by-payment details contained in personal records may eliminate the summary story-telling, yielding more true changes off the seam and fewer false changes on the seam. Furthermore, the new focus on record use may yield more accurate dating of income receipt.

Other procedural changes may also have contributed to the seam bias reduction. For example, the overlapping reference periods (used in the Canadian Labour Market Activity Survey (Murray et al., 1991)), resolution of income source discrepancies across interviews, and the resolution of differences in dates and amounts of income received during the overlap period, were all included in the procedures as a direct effort to reduce the seam bias.

The seam bias result is preliminary and is based on only the small number of cases in the first pretest (n=74). The analysis will be repeated for the data from the second pretest and in the Evaluation Study. In the meantime, the early indication is that the revised procedures were quite successful at reducing, and maybe even eliminating, the seam bias.

C. Areas Needing Improvement - Household Nonresponse and Costs

Although this was not the focus in the pretests, we did compare the pretest nonresponse rates and costs to those for regular SIPP. Both rates and costs appeared higher in the pretests. Table 5 shows the pretest response rates, calculated as the number of interviewed households divided by the number of eligible households. For Wave 1, the response rate was 73%. The Wave 2 response rate, based on the number of eligible cases in Wave 2, was 87%. Overall, the longitudinal response rate was 63%. This rate represents the number of households that were interviewed in both waves based on the number eligible in the first wave. Regular SIPP reports a Wave 1 response rate of about 92% and a cumulative response rate at Wave 2 of about 88%. These numbers are not directly comparable to the rates from the first pretest because there are some important differences in the procedures (such as following movers in regular SIPP). However, it is quite clear that the pretest response rates were much lower.

Likewise, 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 first pretest had higher field costs, perhaps as much as 50% higher.[3]

We examined information captured in the field for evidence that the new procedures caused the high costs and nonresponse. The interviewers described the circumstances of each noninterview they encountered; there is scant evidence in these reports that people refused because of the procedures that were being tested. People did not refuse because they were asked to get records or because they were going to be tape recorded. Most respondents refused immediately, before the interviewer could even begin to explain the purpose of the survey and what was involved. Unfortunately, potential refusals were not always identified soon enough to take any corrective action, and when they were, there was not always immediate followup.

With regard to the effects of the new procedures on costs, one hypothesis, of course, is that such procedures--maximum self-response, group interviews, an appropriate interview setting, the use of records, etc.--required many more visits to the households and, thus, additional costs. The records of visits

[3]Field Division information shows that regular SIPP costs about $37 per case and SIPP-CR is about $54 per case.

from Wave 1 of the first pretest do not show an unreasonable number of "extra" household visits (visits we subjectively judged would not have been necessary under standard SIPP procedures) in order to comply with the survey goals. The information concerning the reason for the contact was often incomplete, which does not allow for a precise count of the number of extra calls and visits. We estimate that between 5% and 14% of all Wave 1 personal visits to interviewed households were extra, depending on how strict a definition is used. There were many extra telephone calls made to households in order to get amounts and dates from records that were not available at the time of the interview. However, it is unlikely that these extra phone calls could explain the higher costs. (The call record data from the second pretest have not yet been analyzed.)

A more likely cause of the higher costs was the fact that the interviewers made many unproductive household visits (Krasko, 1992). There was a clear avoidance of interviewing in the evenings, so interviewers made repeated daytime visits that did not yield productive interviews. Since travel costs are a large part of field costs, it is likely that these non-productive visits contributed to the higher direct interviewing costs. It is possible that the interviewers' inexperience (limited experience on the decennial census and little or no experience 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.

Another possible contributor to higher costs is that the actual in-house interviewing time was somewhat higher for the revised procedures. For the second pretest of SIPP-CR, a Wave 1 interview took an average of 71 minutes per household (2.1 persons per household). For regular SIPP, the average is about 52 minutes per household (2.1 persons per household). It is unlikely, however, that this difference explains all of the much higher cost of the SIPP-CR interview.

We expect that additional training and closer supervision will yield improvements in the areas of nonresponse and cost. It is a common assumption that response rates improve as interviewers gain experience. In addition, the Field Division at the Census Bureau has made some recommendations that are intended to improve response rates and costs, and these will be implemented for the Evaluation Study (Krasko, 1992). Namely, evening and personal visits early in the month will be emphasized, schedules for flow of work will be enforced, staffing decisions will be made to give the supervisor more room for turnover, some interviewers will be hired from within the sample neighborhoods, and interviewers will be provided monthly feedback about their adherence to the quality-oriented procedures as well as about their response rates and costs per case.

VI. PLANS FOR THE EVALUATION STUDY COMPONENT OF THE SIPP-CR PROJECT

The Evaluation Study is planned for September 1992 through April 1993 in Milwaukee. It will include two waves of interviewing and will use a full four-month reference period for both waves. There will be experimental and control treatments: half the cases will be conducted under the redesigned procedures and half will be completed using standard SIPP procedures and instruments. The sample size is 350 completed cases per treatment. Sample names for both treatments will be selected from one of five administrative

files: Aid to Families With Dependent Children (AFDC), Food Stamps, Supplemental Security Income (SSI), Unemployment Compensation, and a Milwaukee employer.

The purpose of the Evaluation Study is to provide a direct comparison of measurement quality across the two treatments, using administrative records as the criteria for assessing quality. Program participation and amounts as reported by the respondents will be compared to the "true" information in the administrative files. In addition, cost component comparisons (travel time, interview time, edit time, etc.) will be made across the two treatments to evaluate the cost of the new procedures and to identify the cause of any additional costs. Lastly, in addition to a straightforward comparison of nonresponse rates across treatments, we will use the administrative data to compare characteristics of nonrespondents across treatments. This will give us some indication of differences in nonresponse bias across treatments.

If there are sufficient quality improvements with the new procedures, with reasonable costs and reasonable nonresponse, further research will be conducted to answer the many operational questions (generalizability to other sites, cooperation over multiple waves, use of computer assisted personal or telephone interviewing, differential effects on subgroups, effect of the different components of the new procedures, etc.) that cannot be answered with the limited research design of the Evaluation Study.

REFERENCES

Burkhead, D. and J. Coder (1985). "Gross Changes in Income Recipiency from the Survey of Income and Program Participation." Proceedings of the Social Statistics Section, American Statistical Association, Washington, D.C.

Krasko, N. (1992). "SIPP-CR Pretest I Type A Analysis." Unpublished 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 Annual Research Conference, Census Bureau, Washington, D.C., pp. 721-745. Also available as Report No. 9008 in the 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. 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, Census Bureau, Washington, D.C., pp. 715-758.

Singh, R. (1991). "SIPP 91: Wave 1 Results of the Record Check Study." Unpublished 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 memorandum for the SIPP Research and Evaluation Steering Committee, June 15, 1992.

REFERENCES FOR ADDITIONAL INFORMATION ABOUT SIPP MEASUREMENT ERROR RESEARCH

Jabine, T. (1990). "SIPP Quality Profile, U.S. Bureau of the Census.

Marquis, K. and J. Moore (1989). "Response Errors in the Survey of Income and Program Participation: Preliminary Results." Proceedings of the Annual Research Conference, Census Bureau, Washington D.C., pp. 515-536.

Marquis, K., J. Moore, and V. Huggins (1990). "Implications of SIPP Record Check Results for Measurement Principles and Practice." Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 564-569.

Moore, J. (1990). "Proxy Reports: Results from a Record Check Study." Paper presented at the International Conference on Measurement Errors in Surveys, Tucson, AZ, November 1990.

Moore, J. and K. Marquis (1988). "Response Errors in SIPP: Some Beginning Estimates Based on Administrative Records." Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 360-365.

Figure 1. Definition of the Seam

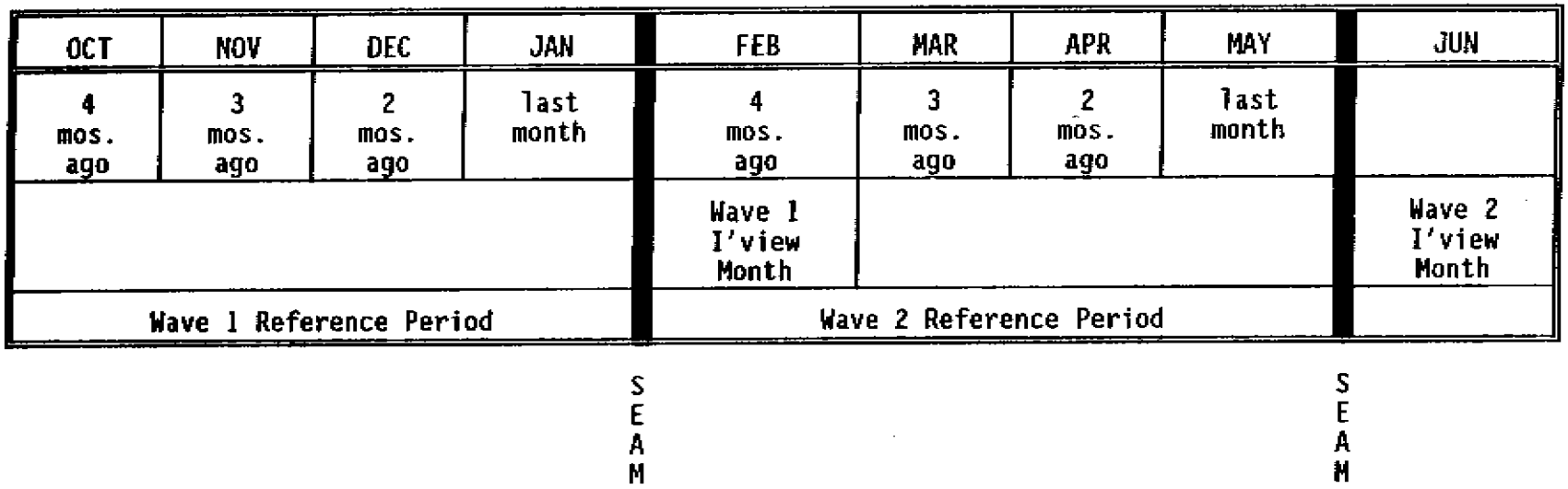

Figure 2. Free Recall of Income Question

This survey is about the economic situation of people in the United States. Our goal is to get a complete, accurate list of all the income, pay, and other money that you (both/all) received during the last 4 months.

IF MORE THAN 1 ADULT:
This includes income for *(read names to confirm all adults covered in current interview, including persons for whom proxy information is being given).*

(Show Payment-by-Payment worksheet) **Here is the worksheet I'll be using. I need to list who received the money, where it came from, the date it was paid, and the amount before deductions. Accuracy is very important so please use all the records you have for each kind of income.**

(Hand calendar to each adult present.) **Here is a calendar. The time period is the past 4 months** *(name months)* **and up to today,** *(read month and date).*

What pay and other money did (you/each of you) get since _____ 1st?

Figure 3. Worksheet Used to Record Income Information

PAYMENT-BY-PAYMENT WORKSHEET

PGM 33
CONTROL No. 300/_____/__/__/___ WAVE _1_ DATE OF INTERVIEW __ __/__ __ 19__ __ SHEET ____ OF _____

ISL CODE _____ FOR SOURCE Mentioned in Question No. _____

PGM 341

RECIPIENT(S) (Person Nos.)	ISL 1,2,8,20-24,27 Who is covered?	SOURCE OF INCOME (Word Descr.)	WHAT RECORDS COME WITH THAT SOURCE?	DATE (MM/DD/YY)	AMOUNT BEFORE DEDUCTIONS	RECORDS USED? YES NO	CHECKLIST (ISL 1-38, 50 or 54?)
a. _ _ _	Recipient covered? 1 [ ] Yes 2 [ ] No		(Mark all that apply) --> PGM 342	1 [ ] No income rec'd from this source during ref. period			1. (Ask or Verify): Were you (getting/covered by) (source) on ___1st? 1[ ] Yes - Ask 2 2[ ] No - Go to 4
			1[ ] Pay statement or check stub	i.	.	1[ ] 2[ ]	
			2[ ] Bank account statement	j.	.	1[ ] 2[ ]	
			3[ ] Asset account statement	k.	.	1[ ] 2[ ]	
			4[ ] Passbook	l.	.	1[ ] 2[ ]	2. Working back from ___1st, when did you last start (getting/ being covered by) (source)? ____/____/____ Month Day Year
	Others covered?		5[ ] Authorization letter	m.	.	1[ ] 2[ ]	
b. _ _ _	d. _ _ _		6[ ] Other (specify)	n.	.	1[ ] 2[ ]	
				o.	.	1[ ] 2[ ]	
			9[ ] None --> PGM 343	p.	.	1[ ] 2[ ]	
c. _ _ _	e. _ _ _			q.	.	1[ ] 2[ ]	3. Did it stop between then and ___1st?
		HOW OFTEN RECEIVED?	WHAT RECORDS WERE USED?	r.	.	1[ ] 2[ ]	
				s.	.	1[ ] 2[ ]	[ ] Yes - 3b. When did it last start again? (Put correct date in 2)
		1 [ ] Weekly	(Mark all that apply)	t.	.	1[ ] 2[ ]	
	f. _ _ _	2 [ ] Every 2 weeks	1[ ] Pay statement or check stub	u.	.	1[ ] 2[ ]	[ ] No - Go to 4
		3 [ ] Twice a month	2[ ] Bank account statement	v.	.	1[ ] 2[ ]	
		4 [ ] Monthly	3[ ] Asset account statement --> PGM 344	w.	.	1[ ] 2[ ]	
		5 [ ] Quarterly	4[ ] Passbook	x.	.	1[ ] 2[ ]	4. [redacted] Any missing records?
		6 [ ] Twice a year	5[ ] Authorization letter	y.	.	1[ ] 2[ ]	[ ] Yes a. Request from source b. Give record-keeping instructions
		7 [ ] Annually	6[ ] Personal notes or ledger	z.	.	1[ ] 2[ ]	
	g. _ _ _	8 [ ] Other (specify)	7[ ] Other (specify)	aa.	.	1[ ] 2[ ]	[ ] No
OFFICE USE ONLY 1 [ ] Yes 2 [ ] No				bb.	.	1[ ] 2[ ]	
	h. _ _ _		9[ ] None	cc.	.	1[ ] 2[ ]	5. [ ] GIVE FEEDBACK

NEXT WAVE REVIEW 1 [] Non-matched source

Table 1. Pretest 1 Self-Response and Group Interview Rates

- 92% Wave 1 self-response (of all interviewed adults, n=168)
  (vs. ≈65% for standard SIPP)

- 75% Wave 1 group interviews (of all interviewed adults in multi-adult households, n=143)
  (N/A for standard SIPP; all individual interviews)

Table 2. Tape Recording Rates

	% of all Interviews Tape Recorded
Pretest 1	73% (n=166)
Pretest 2	77% (n=167)
TOTAL	75% (n=333)

Table 3. Record Use Results

	PRETESTS 1 AND 2			REGULAR SIPP
	Wave 1	Wave 2	TOTAL	(Wave 1)
Percent of **Households** using at least one record	87% (n=178)	88% (n=152)	87% (n=330)	NA
Percent of **Income Sources** reported using at least one record	67% (n=867)	78% (n=714)	72% (n=1581)	20%
Percent of **Individual Payments** reported using a record	57% (n=4229*)	74% (n=2116*)	63% (n=6345)	NA

* The number of individual payments at Wave 2 is reduced from Wave 1 largely because the Wave 2 reference period is half the length.

Table 4. Seam Bias Results

- collapsed across all income types
- Pretest 1 only
- Number of "seam" changes 19
  Average number of "off-seam" changes
  (n=80 over 4 off-seam pairs of months) 20.0
  SEAM BIAS INDEX* = 0.95

compare to

--> *Seam Bias Index for standard SIPP (Burkhead and Coder, 1985):*

Unemp. Comp.	*1.9*
Earnings	*2.2*
Food Stamps	*3.5*
Social Security	*3.9*
AFDC	*4.9*
Priv. Pensions	*6.3*

* The Seam Bias Index is a ratio of the average number of month-to-month changes on the seam to the average number off the seam. It is an index to equalize the fact that there are more off-seam pairs of months then on-seam pairs of months.

Table 5. Response Rates for SIPP-CR

	Wave 1	Wave 2	% of Wave 1 eligible HHs interviewed in both waves
Pretest 1			
Interviews	92	74	
Eligibles	119	89*	
Response rate	.77	.83	.64
Pretest 2**			
Interviews	88	79	
Eligibles	127	86*	
Response rate	.69	.92	.63
Overall			
Interviews	180	153	
Eligibles	246	175*	
Response rate	.73	.87	.63

* The number of Wave 2 eligibles is not equal to the number of Wave 1 interviews because 3 cases in Pretest 1 and 2 cases in Pretest 2 became ineligible at Wave 2.

** Although the interviewers for the two pretests were the same, the samples for the two pretests were very different. The second pretest sample addresses were selected from 5 administrative files, and as we discovered problems with the addresses, we started looking for the sample persons and not the addresses. For this reason, it is inappropriate to compare the response rates across pretests.