CAN WE GET RESPONDENTS TO USE THEIR PERSONAL INCOME RECORDS?

Karen Bogen, Jeffrey C. Moore, and Kent H. Marquis, U.S. Bureau of the Census[1]
Karen Bogen, Center for Survey Methods Research, Census Bureau, Washington, D.C. 20233-4700

KEY WORDS: Income, Record use, Response rate

1. INTRODUCTION AND BACKGROUND

The Survey of Income and Program Participation (SIPP) is one of the U.S. Census Bureau's major, continuing demographic surveys. SIPP provides extensive information on the economic situation of persons and families in the United States. Recently, the SIPP program fielded an experiment comparing standard field procedures to a new way of obtaining the same information about family income and program participation. This paper focuses on one aspect of the alternative procedures, use of personal income records.

Previous research has revealed important measurement error problems in SIPP (Marquis and Moore, 1990 and Marquis, 1990). To gain insights into the causes of these errors, the Census Bureau implemented a small, exploratory cognitive research project to look for clues about respondents' difficulties understanding SIPP tasks and questions, and their thought processes while answering the questions (Marquis, 1990). This research provided important insights into the likely causes of SIPP's response error problems, and led directly to many of the features of the experimental measurement procedures.

The cornerstone of the new measurement procedures was an emphasis on respondents' use of personal income records to assist income reporting. This was a direct result of observing respondents' tendency to use simple, short-cut strategies for "recalling" their 4-month income history. Results of the current research show that interviewers using the experimental procedures were fairly successful at getting respondents to use their income records, but there may have been some costs. The standard approach achieved high response rates, while response rates in the experimental procedures were relatively low. In addition, the field costs for the experimental treatment were much higher than for the standard treatment. This paper describes the record use procedures for the experimental treatment, reports interviewers' success at implementing those procedures, and provides some ideas about the costs or consequences of the procedures.

2. THE EXPERIMENT

The SIPP Cognitive Research Evaluation Study was conducted in Milwaukee, Wisconsin. It included two waves of interviewing, 4 months apart, each with a four-month reference period. Sample cases consisted of individuals, and their associated household members, drawn from the record systems of one of five income sources. We completed between 600 and 700 initial interviews (Wave 1) and between 350 and 400 followup interviews (Wave 2) in households that were randomly assigned to one of two treatments: the standard SIPP measurement procedures (the control treatment) and the redesigned procedures (the experimental treatment). The main purpose of the Evaluation Study was to directly compare measurement quality across the two treatments, using administrative and employer records as the primary criteria for assessing quality. (See Marquis, Moore, and Bogen, 1994, for a discussion of data quality results from this experiment.)

3. THE RECORD USE PROCEDURES

Interviewers in the experimental procedures asked respondents to refer to personal income records, such as pay stubs, bank statements, receipts that come with certain payments, or a personal ledger, to report income dates and amounts. Following are the ways in which record use was emphasized and encouraged in the redesigned procedures:

- The idea of personal record use was introduced as the norm, the usual procedure for reporting a detailed, exact, 4-month income history. Interviewers suggested record use as if it were an expected part of the respondent's task.
- Interviewers asked respondents to try to replace missing records. For example, if the respondent reported having thrown away his last pay stub, the interviewer was to ask the respondent to ask his employer for another copy.
- Interviewers were instructed to make return calls or visits for missing records. If a respondent had used records for other payments of the same type, for example, for 3 of his 4 paychecks, but was missing one, the interviewer could make a phone call to get the last amount from an income record. However, if at the initial personal visit, the respondent did not use any records for a particular source, the

interviewer was supposed to make a personal visit to obtain the reports from records if the respondent agreed. The purpose of this was to make sure the respondent was reading the record information correctly.

- Interviewers provided training for record use in the next wave of interviewing. Interviewers instructed respondents to save all paystubs, bank statements, and any other records indicating the amounts of money received. They provided record-keeping training that was targeted and appropriate for each specific income source. For example, for income from jobs, respondents were trained to keep the pay stubs. For payments from government and public assistance programs, respondents were trained to write down the amounts because the payments often don't come with receipts.
- Interviewers gave respondents a folder in which to save their records, including a special form on which to describe income not accompanied by any records.
- Record use was covered extensively in the initial interviewer training and in the interviewer's manual. For example, in the sections describing the major government programs, we described records that recipients typically get. We also provided guidance on how to ask for records and how to handle any respondent reluctance to use records.
- Interviewers asked respondents' permission to call to remind them to keep records. The reminder call was supposed to coincide with the household's receipt of an important source of income.
- During the interview, the new procedures collected the exact dates of receipt and individual, "to-the-penny" income payments, not monthly totals as are collected in the standard SIPP procedures. The rationale for this was that it would make it easier for respondents to use and interpret their records, and, at the same time, would impress upon both interviewers and respondents the importance of accuracy.
- Interviewers were trained to accept "inefficient" time uses in order to let respondents get records. This is counter to the more typical efforts to get interviewers into and out of a household as quickly as possible. Interviewers were instructed to simply wait if respondents were willing to get records, even if the respondent said it could take awhile to find them.
- The new procedures insisted that the first interview be self-response because the respondent is most likely to be able to locate his or her own records. In addition, interviewers tried to arrange "family-style" interviews, all eligible adults together, since it gave implicit permission for later proxy use of records.
- The final major component of the record use procedures was the use of tape monitoring to provide interviewers with feedback about their quality-oriented performance. This included measures of interviewers' effort at getting respondents to use records and training respondents for future record use.

There were no specific record use "procedures" for the control treatment. The regular SIPP interviewer's manual instructs interviewers to allow the respondent time to locate records, yet an important component of the interviewer's reward and feedback structure is an evaluation of efficiency. In fact, control treatment interviewers were explicitly told in training that they would not be evaluated on respondents' use of records, and were instructed to accept an estimated amount if records were not available.

4. IMPLEMENTATION RESULTS

Standard SIPP procedures capture record use through interviewers' reports about whether any records were used for each reported income source. For the experimental procedures, we assessed record use both through interview monitoring and through interviewers' reports. Furthermore, experimental treatment interviewers' reports were recorded in two ways. First, for each income source, interviewers used a summary item to indicate whether any records were used for that income source (much like the standard treatment), and if so, what kinds of records were used. In addition, as they recorded the date and amount of each individual payment from an income source, interviewers also indicated whether a record was used in reporting that specific payment.

This latter measure permits an analysis of record use at the individual payment level (only in the experimental treatment, of course -- standard SIPP does not capture individual payment information). Both types of interviewer reports, as well as the monitoring results, are shown in the tables below to compare record use rates between treatments at the income source and household levels. In general, the results are quite consistent, regardless of the measure used to estimate record use in the experimental treatment.

4.1 Record Use

Table 1 shows, by treatment, the percentage of households which used at least one record. Based on the interviewers' reports, at least one respondent in about 71 to 74% of all Wave 1 experimental households used at least one record to report about any income source. The monitoring estimate, 68%, is slightly lower, but clearly in the same realm. For the control treatment using the standard procedures, the Wave 1 household record use rate was significantly lower at 25%.[2]

TABLE 1: PERCENT OF HOUSEHOLDS USING ANY RECORDS

	Experimental Treatment			Control Treatment
	Interviewer Reports		Monitoring	
	Payment Lines	Summary Item		
Wave 1	71% (n=609)	74% (n=609)	68% (n=180)	25% (n=700)
Wave 2	84% (n=366)	87% (n=366)	87% (n=137)	22% (n=404)

At Wave 2 for the experimental treatment, the interviewers and the monitors consistently report that about 84-87% of all households used at least one record at the second interview.[3] For the control treatment, the rate was significantly lower than the experimental treatment at 22%, based on interviewers' reports.[4]

Table 2 looks at record use at the income source level, showing the percentage of all income sources that were reported with the assistance of at least one income record. According to the interviewers' reports, about half of all Wave 1 experimental treatment income sources were reported using at least one record. (The monitoring estimate is a little lower -- 42%.) The Wave 2 data show a significant increase to about two-thirds of all sources having their amounts reported with at least one record.[5]

The source-level record use rate for Wave 1 was only 12% for the control group, and it was 11% at Wave 2.[6] This record use rate for the control group is a little lower than regular SIPP achieves, which is about 20% in Waves 1 and 2 (Singh, 1991 and Singh, 1992).

Table 3 shows a third way to look at the record use rates -- the percentage of all of the individual payments that respondents reported that were supported with records. Here, the estimates from interviewers' reports and from the monitoring are identical at 39%.

TABLE 2: PERCENT OF INCOME SOURCES REPORTED WITH ANY RECORDS

	Experimental Treatment			Control Treatment
	Interviewer Reports		Monitoring	
	Payment Lines	Summary Item		
Wave 1	49% (n=2,343)	51% (n=2,343)	42% (n=821)	12% (n=3,004)
Wave 2	69% (n=1,481)	70% (n=1,481)	63% (n=615)	11% (n=1,716)

At Wave 2, over 60% of all individual payments were reported with records.[7] (There is no comparable number for the control treatment, since standard SIPP procedures collect record use information only at the source level.)

TABLE 3: PAYMENT LEVEL RECORD USE

	From Worksheet Payment Lines	From Monitoring
Wave 1	39% (n=12,384)	39% (n=3,758)
Wave 2	63% (n=7,749)	63% (n=2,998)

4.2. Interviewer Behaviors and Perceptions

The experimental procedures included a variety of new field practices intended to help ensure maximum use of records by respondents. Below we summarize results indicating the extent to which interviewers in the experimental treatment actually followed these practices. In general, the results regarding interviewers' acceptance of these novel procedures are less impressive than the "bottom line" results concerning record use.

For example, interviewers were supposed to ask respondents to replace missing records. Monitoring results indicate that this was done in only about one-fourth of the cases where there were one or more missing records. Likewise, we know from monitoring that interviewers were not very good about scheduling callbacks for missing records. Only about 11% of the time did interviewers do what they had been instructed to do -- arrange a specific callback time to obtain the information from the missing records.

Another important new procedure was that interviewers were supposed to train respondents to keep records for the next wave of interviewing. Some of the training was supposed to be source-specific, describing to the respondent exactly how he or she should keep records for this source for the next interview. We know from the monitoring that this happened in only about 40% of the cases when it was supposed to happen, during the interview at the time that income source was being discussed. (However, interviewers were quite a bit better about reviewing source-specific record-keeping at the end of the interview, where the number was closer to two-thirds.) Interviewers gave general record-keeping instructions for possible new income sources only about half the time.

Substantial efforts were supposed to be directed towards changing the habits of non-record-keepers. Interviewers were supposed to ask all non-record keepers to keep records for wave 2 and were to ask them if they would accept a reminder phone call. Monitoring indicated that procedures for arranging reminder phone calls was followed only about two-fifths of the time. Furthermore, based on interviewers' comments, and on the fact that they returned only a couple of forms indicating they had made these calls, we are fairly sure that even fewer reminder calls were actually made.

A critical new procedure was the use of tape monitoring to provide interviewers with feedback about their quality-oriented performance, including persuading respondents to use records and training respondents to save records for future interviews. It was essential to the experimental procedures that interviewers understand that the priorities of this data collection were not the same as for standard SIPP or most Census Bureau surveys. Interviewers were to be given feedback about the things that we believed were important in order to achieve the highest quality data. Much less emphasis was to be placed on the more typical feedback in which the priorities are high response rate and high productivity.

There were problems conveying this message, especially at the beginning of the field work. It took more time than we anticipated to get the whole feedback system into place. Interviewers were not given quality-oriented feedback as quickly as would have been ideal. In addition, the experimental treatment suffered from extremely low response rates, right from the start. As a result, the regional office staff pushed the interviewing staff very hard in order to boost those rates, and thus the "quality" message was diluted.

Interviewers' responses to a debriefing questionnaire sent out in the last month of wave 1 interviewing may provide evidence of this compromised data quality priority. Interviewers were asked the following question:

> Assume:
>
> a. You want to get the highest job rating possible from the RO Supervisor, and
>
> b. You have "10 units" of effort to spend trying to meet the 3 main goals.
>
> How much of your 10 units of effort would you spend on each goal? (Enter a number beside each goal. The numbers should add to 10.)
>
> _____ Keeping costs down
> _____ Getting accurate responses
> _____ Getting a high response rate
> 10

We expected that the interviewers in the experimental treatment, whom, we believed, had been inundated with the "quality" message, would allot the most units to "getting accurate responses". However, the experimental interviewers actually placed that second, behind response rates. It was the control treatment interviewers who selected quality as their highest priority. These ratings suggest that the experimental interviewers heard the message that they were doing poorly on response rates, and that the "quality" message was muffled by this competitor.

4.3. Response Rates and Costs

Table 4 shows that the standard SIPP procedures achieved high response rates, while the experimental procedures got relatively low response rates. In addition, the table shows that the field costs on the experimental treatment were much higher than on the standard field procedures for SIPP.

TABLE 4: EVALUATION STUDY RESPONSE RATES AND WAVE 1 FIELD COSTS

	Experimental Treatment	Control Treatment
Wave 1 Response Rate	83%	95%
Wave 2 Response Rate	90%	98%
Wave 1 Interviewer Costs per Case	$51	$24

5. DISCUSSION

The question posed in the title of this paper is "Can we get respondents to use their personal income records?" The record use rates in the experimental treatment suggest that the answer is "Yes," to a surprising extent. While not every respondent used records for every income source, we saw much more records-based reporting, even at the first interview, than one would expect based on the typical response to a description of our procedures, which was "I would never let you see my personal records."

We also saw, however, that high rates of record use are associated with increased nonresponse and increased field costs. The question is were these low response rates and high costs caused by the experimental procedures, and, in particular, the emphasis on using records? We don't know the answer with any certainty, but there is evidence of some relationship between the two.

If there is a causal relationship, it is not a simple and direct one. Almost without exception, the refusals in the experimental treatment occurred on the doorstep, very early in the interaction between interviewer and respondent. Respondents could hardly have been objecting to the procedures, since they did not know yet what they were; interviewers, however, did know what the procedures were, and we consider the effect of the procedures on them.

We attempted to make all of our procedures completely consistent in conveying the message that data quality was critical and that sloppy, estimated reports were unacceptable. The research design does not allow us to assess the individual effects of different procedures, such as gains from reminder calls versus the request for replacement records. However, we wonder if we went a little too far in our demands on the interviewers; some interviewers certainly felt that they were imposing burdensome demands on the respondents. Perhaps the effects of both these kinds of demands -- those we placed on the interviewers and those that interviewers placed on respondents -- was to make the interviewers less willing and able to sell the survey on the doorstep. In addition, the experimental procedures may have led to some small part of the higher costs, since the record-use procedures required additional visits.

The final result is that the experimental interviewers were quite successful at getting respondents to use their income records, despite the fact that interviewers were not all that compliant with some of the procedures. This suggests possible compromises for the next test of procedures that stress record use. For example, reminder calls and requests for replacement records, which interviewers hardly ever carried out, might be dropped with few consequences for record use but with positive consequences with regard to the perceived burden of the interview. Other possibilities for streamlining the procedures should also be pursued. The goal is to find a way to make the procedures less burdensome and to reduce interviewers' perception that they are placing great demands on respondents by following the record use procedures, and, at the same time, maintain clear commitment to record use.

References

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 (1990), "Measurement Errors in SIPP Program Reports." Proceedings of the U.S. Bureau of the Census' 1990 Annual Research Conference, March 18-21, 1990, pp. 721-745.

Marquis, K., J. Moore, and K. Bogen (1994), "An Experiment to Reduce Measurement Error in the SIPP: Preliminary Results." Presentation at the August 1994 American Statistical Association Conference (this volume).

Singh, R., "SIPP 91: Wave 1 Results of the Record Check Study." Internal Census Bureau Memorandum, December 19, 1991.

Singh, R., "SIPP 91: Wave 2 Results of the Record Check Study." Internal Census Bureau Memorandum, June 15, 1992.

Notes

[1] The authors thank Yuki Ellis of the Census Bureau's Center for Survey Methods Research for her important contribution to the data analysis. The views expressed are attributable to the authors and do not necessarily represent those of the U.S. Census Bureau.

[2] T-test comparisons of household record use in the experimental treatment vs. the control treatment were significant for both the rate based on payment lines and the rate based on the summary item ($p < .05$).

[3] Based on just the 351 experimental treatment households that participated in both waves, there was significant improvement in record use from Wave 1 to Wave 2 ($t=3.35$, $p<.05$ based on payment lines; $t=3.69$, $p<.05$ based on the summary item).

[4] $p<.05$ using t-test comparison of either Wave 2 payment lines (experimental treatment) vs. control treatment or Wave 2 summary item (experimental treatment) vs. control treatment.

Also, for the control treatment, there was a significant decrease in household record use from Wave 1 to Wave 2 for the 376 households that were in both waves ($t=-1.97$, $p \approx .05$, two-tailed).

[5] Again, looking at just the experimental treatment respondents who were in both waves, there was significant improvement from Wave 1 to Wave 2 in the average proportion of sources reported using records ($t=7.39$, $p<.05$ based on payment lines; $t=7.68$, $p<.05$ based on the summary item).

[6] The control treatment source-level record use rates are significantly lower than for the experimental treatment in both waves ($p<.05$ for all comparisons between experimental and control treatments).

[7] For persons who were in both waves, there was significant improvement from Wave 1 to Wave 2 in the average proportion of individual payment amounts reported using records ($t=11.04$, $p<.05$).