THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

RECORD USE BY RESPONDENTS

No. 152

R. Kominski
Bureau of the Census

U.S Department of Commerce BUREAU OF THE CENSUS

RECORD USE BY RESPONDENTS

Introduction

This is about a survey in which large amounts of information about the economic situation of households (and the people in them) are requested from respondents. The Survey of Income and Program Participation (SIPP) is a large national household-based longitudinal survey. Households are in sample for a total of 32 months and are visited by an interviewer every four months (for a total of 8 interviews). During each interview, respondents are asked questions about their employment, economic situation and program participation during the preceding four months. A large part of the intent of the survey is to be able to examine the economic and social changes that occur to people over time. As such accurate information in each interview is especially critical, since the independent interviews will eventually comprise one complete 32-month data file. During interviews we strongly encourage respondents to use records of some kind when telling us about their jobs, the programs that belong to, the sources of incomes they have, or the amount of money each of those sources provides. Interviewers are trained to probe for this information, but not to jeaporadize the interview if the respondent does not seem willing of interested in using records. In these cases, interviewers are instructed to take whatever information they can obtain.

In the course of our work on the SIPP over the past few years we have become concerned about the validity of some of the information provided by respondents. The first indication of a problem cam from an administrative record check study of some of the cases where social service program participation wareported. In several papers by Marquis and Moore (1989a, 1989b), the authors concluded that there was a substantial amount of error both in the reporting of specific programs. In simple descriptive terms, underreporting of Social Security to 37% percent for AFDC. Marquis and Moor also showed that the error in reporting "rolled through" into other more analytic program changes. While at least part of this error was attributable to between-interview conditions (the "seam" effect) a larger part of cases the answers we are getting are wrong; either the person does not report a program when in fact they are on one, or reports a program, but does so incorrectly.

A second indication of suspect date comes from an examination of a set of one time "topical module" questions on income tax amounts. Topical modules are sets of questions on a special topic that are generally asked in only one interview. During the 5th interview of each panel in the SIPP, respondents are asked a series of questions about their taxes during the preceding year. These questions are asked after April, that is, after the point in time when everyone should have filled their taxes and have records available to them. For a subset of the questions, the W-2 form will suffice to provide answers; for another subset, the federal tax form will provide the "correct" data. An internal memorandum (Altman, 1989) documented the extent to which respondents when asked to use their W-2 or federal return form, actually did so. Based on the administration of these questions in four different modules (over 2 panels) a W-2 form was used about 38 percent of the time, while the federal tax return form was used in just 37 percent of the cases. Of course, respondents might not have used forms for a number of reasons: they weren't easily available; they couldn't be found after looking for them; or the respondent did not want to

let us see these documents (note that persons who did not respond to the questions at all are not included in these estimates). For whatever reasons, only a third of respondents bothered to use a form that is a fairly important document for most households.

The Study and Data

The results of the administrative record check and the income tax data, along with numerous observations of actual interviews, and a small cognitive research project aimed at understanding respondents' difficulties with questions, led us to believe that records were not used by many respondents. In an attempt to ascertain the basic level of record use, two different research studies were designed. The first, discussed here, uses Senior Field Representatives (SFR's) who did routine observations of Wave 1 interviews in the 1990 panel of SIPP. The second study, begun with the second wave of the 1990 panel, has the Field Representatives (interviewers) check off all income source codes for which Record was used. This is done using the Income Source Summary (ISS) listing on the back page of the SIPP questionnaire. This second study is still underway; preliminary results will be mentioned at the end of this paper.

In every wave of SIPP, a small subsample of all interviewers are chosen to be observed by their supervisors (SFR's). These observations are part of ongoing activities to review the field interviewing staff on a routine basis. Interviewers to be observed are chose fairly randomly, with the limitation that all interviewers are observed twice a year. The primary purpose of these observations is to review things such as form and behavior, and attempts to identify problems that the interviewer may be having.

A form was designed for SFR's to use when conducting observations during Wave 1 interviews. The form allowed the SFR to note whether records were used in the reporting of certain income sources: wages and salary, assets, and certain public programs. Those programs with the largest number of participants nationally chosen for inclusion on the form. SFR's were to note simply if ANY record was used by the respondent to verify an income source. Any kind of record would be acceptable - not only a formal document (program eligibility or payroll notice), but a secondary source such as a checkbook entry or eligibility card. For wages and assets, the SFR would check a yes/no box indicating that a record was or was not used. For the eight programs listed, 3 different boxes were provided for record use; one indicating no record has been used to identify the source only, and a third to indicate a record had been used to verify both the source and amount.

The Wave 1 interview is the first interview of a panel. In many respects this interview sets the tone for all future interviews. If the interviewer can create the right tone and attitude, future cooperation may be assured. On the other hand, a bed start may result in the loss of a household not only for Wave 1, but for the entire panel. While we do not require the use of records for verifying income sources, we strongly urge their use, both to interviewers (in training) and respondents (in the introductory letter, in questionnaire lead-ins, and by interviewers themselves). In addition, since a third party was involved in these particular interviews, we felt that the incentive to "do things by the book" would be even stronger than usual. Thus, we expected some upward bias in terms of the proportion of persons who would use records in some way (were in SFR not present); as such, these estimates might be expected to represent an upper limit in terms

of record use.

Regional offices were instructed to give the Record Use Study form and instructions to all SFR's doing observations during Wave 1. All households visited were to be recorded on the form, with each adult case (persons age 15 and above) recorded on separate lines. SFR's were instructed not to tell interviewers about the form in advance, but, if an interviewer noticed the form and asked about it, the SFR was to tell the interviewer about the study, and note on the form the point at which the interviewer was informed. After observation was completed, the form was mailed back to Suitland for encoding and analysis.

Acceptance of and cooperation with the study were quite good. During the 4-month period of Wave 1, 219 observations were scheduled throughout the country; Record Use Study forms were received for 191 of these. Of those 28 missing, about half were from one region (Charlotte), and a quarter from another (Los Angeles). All other regions were complete or virtually so. These 191 forms recorded information for 466 households containing 891 people, or about 2% sample of all households and people interviewed in Wave 1. Apart from the shortage of cases from the Charlotte Region, there is no reason to expect that this sample is not representative of the full SIPP sample.

Results

Simple frequency distributions of the data show the degree to which records were used by respondents; these are summarized in Table 1. Of the 609 persons reporting a wage or salary, 189, or 31% used a record of some kind. A similar level of use was reported for assets, 171 of 616, or 28%.

The distribution of programs is somewhat less straightforward than with wages or assets. While the sample consists of nearly 900 persons, many of the programs in question are rare enough such that a sample of this size can generate only a small number of participants. One frequently-mentioned program, however, is social security. About a third of the sample reported they received Social Security (271 persons), but 43% of these recipients did not verify this in any way. Of those who did provide verification, 1 in 3 verified that source only. Only about a third of all Social Security recipients (35%) verified both the source and amount with some kind of record. One should keep in mind that even something as simple as a checkbook entry would be counted as a verification of source and amount for this study.

Many persons who receive Social Security also get Medicare, and 147 persons reported this source. A large proportion of these persons (78%) were able to verify Medicare with a record of some sort. Two-thirds of those verifying Medicare did so for the source only; this can be done by showing the interviewer the Medicare card that all recipients get. In fact, given that a section of the SIPP questionnaire asks respondents who report Medicare to show the card to the interviewer, it is somewhat disturbing to see that only 78% of recipients provided verification.

Other programs do not occur in as great a frequency as Social Security, and the observed counts in the sample reflect this. For the other 6 programs, counts of recipiency ranged from 4 (general assistance) to 50 (Medicaid). Rather than look at each of these sources separately, we combined all reports for a single measure. Of the total 147 reports of recipiency, only 31 were verified in some way (21%). About two-thirds of the verifications (22 of 31) were for both the

source and amount, however, these simple results, infrequent, except for Medicare, where an explicit verification (the Medicare card) is always requested.

The next step was to look at factors which might be associated with differential levels of record use. The observation form allowed the SFR to note three important conditions; whether the interviewer was experienced or inexperienced in the survey; if the respondent was a self or proxy; and whether the interviewer knew about the experiment or not. All three of these conditions were thought to influence the level of record use: new interviewers might be thought to be more conforming to the rules of the survey, as would be those who learned of the purpose of the experiment, and proxy respondents might be less knowledgeable and able to locate records.

Table 2 shows the results of simple chi-square tests on the observed data. Wages, assets, social security and Medicare are shown separately; the remaining 6 programs are grouped together. The results in Table 2 indicate that for the most part of the experience of the interviewer had little effect on whether records were used or not. Only for assets did new interviewers yield significantly higher levels of record use. While only a small proportion of cases were done in the condition where the interviewer knew about the experiment (7.7%), there was no apparent effect.

Differences in record use are seen across the self-proxy dimension, but the pattern is not consistent for all items. While wages, assets and Medicare were more often verified with records by self-respondents, social security and the remaining 6 programs were more likely to be verified by proxy respondents. It is important to note, however, that only a small number of programs were reported at all by proxies (just 14 of 144 total for the 6 programs), but that half (7) were verified.

This leads to a more fundamental question; namely, whether the number of reports of recipiency in general (verified or not) varies according to interview factors. Table 3 briefly addresses this issue, concentrating only on the self-proxy part of the problem. The results of the chi-square tests show a strong pattern of higher levels by self reporters. In short, one might infer that self respondents, being most knowledgeable about themselves, generated higher levels of recipiency in their reports, and for the most part, are also more likely to verify them. In some cases, however, proxies may only report recipiency when they are certain about the situation, as when they can verify it through a record of some sort.

While the self-proxy, awareness, and new-old interviewer factors failed to reveal any consistent differences in use of records, there may be some other underlying factor (s) that influence record use. One obvious possibility is that record use may be associated with specific interviewers, households, or respondents. In order to investigate this hypothesis, a datafile of all reported program recipiencies was generated from the original date. In this datafile, each report of any of the eight programs was output a separate case; thus if one person received all eight programs, they would be represented as eight records. These records are then tabulated by specific interviewer, household and person identification numbers, in order to see if record use (or non-use) is clustered within specific units (be they interviewer, household, or person).

The results of the analysis of the record use file for programs are shown in Table 4. There were a total of 565 reported recipiencies among the 8 programs; 266 of these were not verified with a record in any way, and 299 were verified in some way (for purposes of this analysis, we have collapsed verification of source only and source and amount). These 565 recipiencies were

reported over 131 interviewers (out of 191 total, or 69%), 218 households (of 466 total, or 47%), and 318 persons (of 891 total, or 36%). The first panel shows the general results for interviewers, who recorded program recipiencies, 92 (70%) had at least one report with no verification by record; of these, 35 (27%) had no records that were verified. These 35 interviewers accounted for 98 of the total 277 (37%) recipiencies that were not verified.

The table in the second panel partitions the 92 interviewers with recipiency reports into 3 groups: those who had 5 or more verified reports ("bad"); those with 5 or more verified reports ("good"); and those with a "mix" or fewer than 5 of either type (of course, the total number of cases completed by each interviewer influences this number). The fourteen interviewers who had 5 or more nonverified programs obtained verification only about 17% of the time, and accounted for 41% of all nonverified programs, while those in the "high verification" group did so 91% of the time and accounted for just 4% of all nonverified programs. The "mix" group verified 54% of the recipiences they obtained. A chi-square test of this distribution yields a value of 1361.1 with 2 degrees of freedom, confirming that these 3 groups of interviewers had very different levels of record use. Interestingly, both the "good" and "bad" interviewer groups had far more recipiency reports on average (about three times as many) as did the intermediate "mix" group.

The third panel of Table 4 looks at the data from the level of household. In this case, of the 218 households that reported some recipiency, 139 (64%) had at least one report without verification, and 79 of these (36%) had no verification of their reports. These 79 households accounted for 177 of the 266 recipiency reports (67%) where no record was used for verification.

As with households, the table in the third panel shows the distribution of verified and nonverified programs cross-classified by 3 types of households; those with 4 or more nonverified programs cross-classified by 3 types of households; those with 4 or more verified programs, those with 4 or more nonverified programs, and the remainder. The fourteen households with 4 or more nonverified programs had virtually no verifications at all-just 1 of 65 programs (1.5%) was substantiated by a record. These 64 nonverified programs accounted for 24% of all those in the study. By contrast, the 23 households with 4 or more verified programs were virtually complete in doing so - 105 of 108 recipiencies (97%) had a record check, and the 3 nonverifications account for only 1% of all such reports. Verification in the "mix" group was a moderate 49%. The chi-square test statistic of 156.0 (2 d.f.) Shows that the three household types are radically different in their level of verification. As with interviewers, the average number of reports in both "good" and "bad" households was substantially higher (4.6 and 4.7, respectively) than in the "mix" homes (2.2).

The fourth panel of Table 4 looks at record use patterns at the person level. Here, of the 318 persons who reported obtaining at least one of the eight programs, 170 (53%) had at least one unverified source, and 116 of these (36%) had no verification for any source. These 116 people accounted for 201 of the total 266 (76%) recipiencies that were not verified with a record.

The fourth panel of the table also has three groups; persons with 3 or more nonverified programs, 3 or more verified programs, and the remainder, with a mix or fewer than three. The 23 people with 3 or more nonverified programs also had few verifications - just 4 of 82, for 5% verification. In addition, these individuals accounted for 29% of all nonverified programs. The 11 people who had 3 or more verifications had an overall high level - 93% record use, and were responsible for only 1% of all nonverified recipiencies. The "mix" group had verification in 42%

of it's cases. The result of the chi-square test, while somewhat smaller than for interviewers or households (106.8) is still substantial and significant. The average number of programs reported for "bad" (3.6) and "good" (3.7) respondents was about twice as high as for all others (1.6).

This analysis seems to indicate that interviewers and households may be more useful than persons in locating the source of variation in record use. Of course, the chi-square tests are not independent, since bad persons may reside in bad households, and be covered by bad interviewers. The two-way chi-square tests act as rough guides of the ability of each factor (person, household, interviewer) to independently account for the distribution of record use, but cannot be viewed as exact tests, given that the factors are not independent.

As a final check, we examined the specifies of the "best" and "worst" cases we identified at the interviewer, household and person level. In terms of "bad" cases, we had identified 14 interviewers, 14 households, and 23 persons. These 14 interviewers accounted for 11 of the households and 19 of the persons in question; that is, knowing the interviewer alone is useful in identifying most of the bad household or persons. Only 4 of the 14 interviewers were new to the survey, and self and proxy interviews were about evenly mixed. However, most of the programs reported on were those other than social security or medicare.

Examination of the "good" cases showed similar clustering. The 12 interviewers with high levels of record use contained 15 of the 23 good households that were identified, and 9 of the 11 good person cases. Old and new interviewers were about equal in proportion, as were self and proxy interviews. However, unlike the bad cases, a large proportion of the programs reported here were social security or medicare.

Conclusion

The results of the analyses of the record use study indicate that interviewers may be the critical link in getting respondents to use records on a routing basis. While a great deal of variation in record use can be found at the household and person level, many of the worst households and person level, many of the worst households and persons are subsumed by interviewers who have multiple bad households and persons within their workload; conversely, many of the very best persons and households can be identified as belonging to a few very good interviewers.

Examination of these interviewers and the data they collect indicate that record use/nonuse occurs in fairly consistent fashion with interviewers. While factors such as self/proxy and old or new interviewer status bear little on the observed distribution, the specific programs in question do seem important. The best interviewers, with many verifications, had many reports of social security and medicare, while the worst interviewers had many reports of the other programs, which may be more difficult to verify.

Perhaps more interesting are the number of programs each type of situation typically yielded. Regardless of the kind of observation unit (interviewer, household or person), good and bad record use situations yielded average number of recipiencies ranging from 2 to 3 times that obtained in non-extreme ("mix") situations. It is not entirely clear what this means, but one could interpret it to mean that "good" interviewers (or households, or respondents) not only get an extensive portrayal of what goes on, but do so accurately (i.e., with record verification). "Bad"

7

interviewers (household, respondent), on the other hand, manage to get the basic information of recipiency at the cost of any kind of verification on some of the things that are reported in addition to missing some things altogether. If this is in fact the case, we may need to redouble our efforts and direct attention not only at better verification of reports, but on more complete reporting as well.

The fundamental finding, of course, is that across all elements, record use is quite low. Wages and assets are verified less than a third of the time, and only about a fifth of the programs other than social security or medicare were supported by a record of any type. Early data from the Income Source Summary (ISS) experiment (Singh, 1991) show record use levels for amounts in the vicinity of 20%, indicating that the results found here in this study are not unreasonable. The Census Bureau has now embarked on a large-scale cognitive research project attempting to promote record use on a routing basis by respondents and interviewers. An associated record-check study will be used to monitor the quality of the data collected. Hopefully, these research projects will lead the way to an interview situation in which data can more easily and routinely be validated during the collection process.

References

Altman, N.A. (1989). "SIPP Record Usage Rates for the 1984 and 1985 Panels' annual Round-Up Topical Modules." Internal memorandum for Chester E. Bowie, June 30, 1989. U.S. Bureau of the Census.

Marquis, K. and J. Moore (1989a). "Response Errors in SIPP: Preliminary Results." Proceedings of the Fifth Annual Research Conference, pp. 516-536. U.S. Bureau of the Census, Washington, D.C.

Marquis, K. and J. Moore (1989b). "Some Response Errors in SIPP - With Thoughts About Their Effects and Remedies." Proceedings of the survey Research Methods Section, American Statistical Association, Washington, D.C.

Singh, R.P. (1991). "Initial Results of the Record Check Study." Internal memorandum to the SIPP Research and Evaluation Committee, January 10, 1991, U.S. Bureau of the Census.

Table 1.   **Distribution of record use (among recipients only)**

Source	No record used	Source	only	Amount	&	Source
Wage/Salary	414	%	%	%		%
Assets	445			189*		31
Social Security	117	58	21	96		35
Medicare	33	73	50	41		28
AFDC	23	1	4	1		4
SSI	11	2	10	8		38
Food stamps	36	0	0	2		6
Medicaid	34	6	12	10		20
General assistance	3	0	0	1		25
WIC	9	0	0	0		0
Combined 6	116	9	6	22		15

*Verification asked for amounts only

Table 2.   **Record use rates by interview factors**

Source	Interviewer			Respondent			Aware of experiment		
	New	Old	Pr*	Proxy	Self	Pr.	No	Yes	Pr.
Wages	.31	.31	.98	.27	.33	.10	.31	.30	.81
Assets	.34	.26	.01	.19	.31	.01	.28	.20	.26
Soc. Sec.	.64	.53	.08	.61	.56	.50	.57	.53	.74
Medicare	.77	.78	.93	.68	.81	.15	.78	.70	.55
Combined 6	.22	.21	.87	.50	.18	.01	.20	.50	.15

*Significance level of chi-square tests of independence

Table 3.   Level of program reports by respondent type

	Wages	Assets	Soc. Sec	Medicare	Combined 6
Proxy	.75	.62	.20	.10	.01
Self	.65	.73	.35	.20	.04
Signif.*	.00	.00	.00	.00	.00

*Significance level of chi-square tests of independence

Table 4.   Interviewer, Household and Person Patterns of Record Use

	Intv.	Hhld.	Pers.
Total Units	191	466	891
With Programs	131	218	318
Some with no records	92	139	170
All no records	35	177	201
% of programs	98	177	201

INTERVIEWERS:

%	Type	--Verification-- No Record	Record	Average % of Programs Reported	
(14)	"Bad"	109	22	9.4	
(105)	"Mix"	147	173	3.0	Chi-square:
(12)	"Good"	10	104	9.5	136.1

HOUSEHOLDS:

%	Type	--Verification-- No Record	Record	Average % of Programs Reported	
(14)	"Bad"	64	1	4.6	
(181)	"Mix"	199	193	2.2	Chi-square
(23)	"Good"	3	105	4.7	156.0

PERSONS:

%	Type	--Verification-- No Record	Record	Average % of Programs Reported	
(23)	"Bad"	78	4	3.6	
(284)	"Mix"	185	257	1.6	Chi-square
(11)	"Good"	3	38	3.7	106.8