THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

AAPOR ROUNDTABLE: IMPROVING INCOME MEASUREMENT

No. 241

Pat Doyle
U.S. CENSUS BUREAU

U.S. Department of Commerce U.S. CENSUS BUREAU

AAPOR Roundtable: Improving Income Measurement

compiled by Pat Doyle,
U.S. Census Bureau
Demographic Surveys Division,
Washington DC 20233-8400

Key Words: Methods panel, income reporting, data quality, instrument design

Abstract
The Census Bureau established the SIPP Methods Panel project to evaluate and redesign the core instrument for SIPP, a recurring, nationally representative, longitudinal survey of people and their socio-economic characteristics. The objectives of the project are to improve response rates in SIPP, to reduce burden, and to improve data quality. It is a research project consisting of analysis of extant data as well as experimental research. The data analysis component includes examining patterns of nonresponse, examining reporting patterns across waves of interviewing, and analyzing patterns of income receipt. The program of experimental research consists of three phases, designed to allow for three iterations of testing and refining the Wave 1 core instrument and two iterations for Wave 2. Each experiment involves the selection of an independent sample of the population in six regional offices—half of which are to be randomly assigned to a control group, and half assigned to a treatment group. Each sample is designed to have 2000 interviewed cases, with the half in the control group receiving the SIPP instrument in the field at the time of the experiment and the other half in the treatment group receiving an experimental instrument.

The AAPOR roundtable discussion presents findings from the first two field tests conducted in summer 2000 and summer and fall 2001 and uses that to promote dialogue on successful approaches to designing instruments for complex longitudinal surveys. Topics to address include approaches to reducing item nonresponse and income underreporting, methods of collecting earnings and their impact on response rates and data quality, nonresponse follow up techniques and their success, methods of improving interview efficiency and assessment of interviewer satisfaction, the success of new methods of assessing within household coverage, and a cognitive assessment of dependent interviewing techniques.

Background
The Survey of Income and Program Participation (SIPP) is a longitudinal survey conducted by the U.S. Census Bureau to provide data on the distribution of income, wealth, and poverty in the United States, and on the effects of federal and state programs on families and individuals. Currently, SIPP consists of nine waves, or rounds of interviewing, with each wave administered every four months to a nationally representative sample of the civilian noninstitutionalized population. Interviewing for each wave is distributed over four successive calendar months to create a stable production workload for field staff. It is primarily a person-based survey, administering a battery of questions to each person age 15 or older (or their proxy) in interviewed households, using computer assisted personal interviewing techniques (U.S. Census Bureau 2001).

The survey instrument is extremely complex, collecting information about household structure, economic status, income sources, and labor force participation. The current reference period for most questions is the four months before the interview. Core questions are fully administered the first time an individual is interviewed. During subsequent contacts, the instrument uses dependent interviewing techniques to reduce the burden on respondents and to attempt to reduce seam bias effects. ("Seam bias" is said to occur when respondents report month-to-month transitions as occurring much more often *between* survey waves as opposed to between months *within* a single wave. Statistically, such transitions should occur almost evenly across all months of the survey.)

In 1996, the SIPP Executive Committee established the Continuous Instrument Improvement Group (CIIG). Consisting of staff from numerous Census Bureau technical, program, and research areas, the CIIG task was to review the SIPP core instrument—to improve the instrument and, if possible, shorten it to reduce respondent burden. The CIIG generated an extensive set of recommendations, most in need of testing before implementation in the production SIPP instrument. The need for thorough and rigorous testing led CIIG to recommend (and the SIPP Executive Committee to accept)

the creation of a methods panel project, separate from the production survey.

Methods Panel Project

The methods panel project consists of a small research project, conducted in parallel with the production SIPP. The project is experimentally designed to support rigorous testing of alternative SIPP instrumentation. In addition, the methods panel project encompasses reviews of the literature, quantitative analysis of existing and new data, and qualitative analysis of the instrument and the data collection methodology—all with the goal of improving upon the current measurement methods.

The project's primary goals are to improve the quality of SIPP core data—through improvements to individual items and sections of the questionnaire that:

- ✓ Lead to reduced nonresponse to particular survey items.
- ✓ Ease the administration of the instrument by interviewers.
- ✓ Reduce the burden on respondents.

In addition to research and analytic tasks, the project encompasses three formal field experiments, as described in Doyle, Martin, and Moore (2000). Each experiment involves the selection of an independent sample of the population in six regional offices—half of which are to be randomly assigned to a control group, and half assigned to a treatment group. The experiments conducted in summer 2000 and 2001 form the basis of this paper. In each of those experiments, half the sample (the treatment group) received the experimental instrument described in Doyle, Martin, and Moore (2000). The other half (the control group) received the SIPP Wave 1 instrument for the panel in the field at the time. The project employs a multistage, clustered sample design—as is true for the production survey (U.S. Census Bureau 2001)—however, unlike the production survey, the experiments do not over sample poor areas and only include self-representing Primary Sampling Units.

For each experiment, the design dictates the selection of a sample of approximately 1,350 addresses for each of the treatment and control groups, with the expectation that the initial sample would yield approximately 1,000 households interviewed in each group (total n=2,000). Samples of this size in each treatment are sufficient to identify treatment differences in item nonresponse rates from 3 to 8 percentage points, depending on the universe for each item (Doyle, Martin, and Moore 2000).

Table 1 presents the sample sizes and response rates. Note that we did not achieve our goal of 1000 interviewed households per treatment because the assumptions used in sample selection were that response rates for Wave 1 would be in excess of 90 percent. We did redraw the sample for the third experiment fielded in 2002 and achieved the full 1000 cases per treatment.

Features of the Experimental Instrument

Detailed information on the instrument changes introduced are documented in Chan and Moore (2001), Doyle and Moore (2001), Griffiths (2001), Moore (2001), and Pascale (2001). A summary of major changes introduced.

Demographic Sections: We changed the way in which we enumerate persons in the household in order to reduce within-household undercoverage (Doyle, Martin, and Moore 2000), and we used a topic-based format to collect person-level demographic characteristics (like age, race and sex) more efficiently. In response to concerns from the field on the perceived intrusiveness of the questions, the enumeration method was modified in the 2001 experiment and will be refined a bit more in the 2002 experiment.

Labor Force and Earnings: To improve the precision of the questions on labor force status and types of jobs held, we introduced a new four-part series of questions in the 2000 experiment: businesses owned and their owners at the household level; person-level "self-employment"; work for an employer; and miscellaneous jobs. In the 2001 experiment, we also introduced a different approach to collecting earnings—based on giving the respondent more flexibility in choosing the best method for reporting amounts received (monthly, annually, weekly, biweekly, quarterly, or hourly).

Unearned Income: We introduced screening procedures to effectively target need-tested program questions to households that were potentially eligible to receive such benefits. We also introduced expanded questions targeted to the in-kind benefits offered after the introduction of welfare reform.

Assets: We introduced a three-part approach to asset recipiency: first, determining ownership of Individual Retirement Accounts and other individually held retirement funds; next, ascertaining ownership of a select set of more commonly held asset types; and finally,

capturing ownership of the balance of asset types. In the latter case, detailed questions were only asked if the respondents held the more common types, or indicated (in response to a general question) that they held at least one of these less-commonly held types. Joint ownership questions were expanded to refer to all types of jointly held assets—including accounts jointly held with children, household members other than the spouse, and nonhousehold members. Asset income amount questions were changed to increase flexibility to report the amounts in annual or sub-annual amounts. Finally, we expanded the use of nonresponse follow-up questions to allow reporting of a range on the amount received if the actual amount received was forgotten or not known.

Overall Evaluation of the Experimental Instrument

The effects of the instrument changes on data quality are summarized here and evaluated in detail in Chan and Moore (2001), Doyle and Moore (2001), Griffiths (2001), Moore (2001), and Pascale (2001). Below we compare unit-level and item-level nonresponse patterns, and income and program participation reporting patterns, across the experimental and control groups. In addition, we summarize the results of an interviewer debriefing.

✓ *Unit Nonresponse*: *There were no differences in unit-level response patterns across the experimental and control samples or in the composition of interviewed households across the treatments*.

We had expected a nonresponse rate of about 13 percent for both samples in the 2000 experiment, based on the nonresponse rate for the SIPP 2000 Panel. We experienced a significantly higher nonresponse rate among the treatment group in that test. However, no observed differences were significant across the treatment groups and the nonresponse rates in the 2001 experiment were not significantly different from SIPP 2001 Wave 1 (See Table 1). We were disappointed that the interviewer's expectation of a more efficient interview did not have a positive impact on response rates in Wave 1. However, the sample size is small so we would have had to see at least a 3 percent change in response rates for the difference to be significant. Wave 2 is actually a better test of this, however, since both respondents and interviewers are already familiar with the content and duration of the survey.

✓ *Labor Force Participation: The experimental instrument yielded the same overall labor force participation rate*

However, the distribution of jobs held shifted from wages and salaries to less regular types of employment (see Table A). While this resulted in fewer questions administered for some workers, overall time to complete the module did not decrease, as originally anticipated.

✓ *Item Nonresponse*: *The experimental instrument yielded significantly lower item nonresponse than did the control instrument.*

To estimate item nonresponse for a particular topic, we computed the ratio of the number of questions with missing responses for a given person, divided by the number of questions administered to that person on that topic. So, for example, if eight income amount questions were administered, four of which had missing amounts and four had reported amounts, the fraction would be .5. Table 2 illustrates the average of these fractions over persons (or households) in each sample, within various groups of questions defined by income type.

Overall, the treatment group experienced significantly lower item nonresponse on income amounts than did the control group. This was most pronounced for asset amounts in both experiments (both before and after consideration of the nonresponse follow-up items) and for earnings in the 2001 experiment where we allowed flexibility in reporting amounts. For earnings we achieved a reduction in item nonresponse of over 40 percent. We did see some changes in the detailed components of earnings, which are summarized in Table B.

✓ *Income Reporting: Overall, there was no impact on income reporting—as measured by a comparison of the mean amounts and the proportion of the population with income*.

Household total income and person total earnings did not differ significantly at the .05 level. The proportion of the population with the major types of income (earnings, assets, other) did not change significantly either (detailed results available upon request).

✓ *Interviewer Debriefing Results*: *By the time we implemented the full set of changes to the instrument in the 2001 experiment, the interviewers showed a clear preference for the test instrument—with two exceptions.*

The exceptions were the changes in within-household enumeration (which we continue to refine) and the Labor Force 1 which we have since returned to its original structure. Table 3 summarizes interviewer preferences by module and by experiment.

Recommendations for Wave 1

The methods panel team presented its recommendations for Wave 1 of the 2004 production survey to the SIPP Executive Committee in February, 2002. These recommendations covered most components of Wave 1, with a few components needing further testing. These are the highlights of the recommendations:

✓ Adopt general changes to upgrade the instrument to the latest screen standards and CASES options.
✓ Adopt the new roster procedures to collect multiple names and allow more information to be collected than just name (e.g. sex, relationship) as the initial household composition is determined.
✓ Refine the roster probes and try again in the third experiment.
✓ Adopt the new topic-based approach to the collection of demographic characteristics.
✓ Do not implement the new Labor Force 1 approach at this time, planning further study of the results.
✓ Implement the new approach to collecting earnings.
✓ Adopt general household income screening procedures for collecting need-tested benefits.
✓ Adopt new screener procedures for collecting asset recipiency.
✓ Adopt asset nonresponse follow-up procedures.

Conclusion

We successfully implemented a revised SIPP Wave 1 instrument—incorporating several changes designed to improve data quality, reduce nonresponse, or increase the ease with which the interview could be conducted. This experimental instrument was implemented to a sample of households in the summer of 2000, alongside a sample of equal size and design administered the unaltered control instrument for 2000 SIPP Wave 1. We repeated this feat again in the summer of 2001, following that with a revised Wave 2 instrument in the fall of 2001. The results were encouraging. We observed a reduction in item-level nonresponse rates among earnings and asset amounts, and we observed improvements in administration of the survey. Once the planned changes are fully implemented, we are optimistic there will be further improvements in data quality.

References

Chan, Anna Y., and Jeffrey C. Moore. 2001. New Roster Probes and Procedures to Improve Coverage in Household Surveys: The Case of the Survey of Income and Program Participation (SIPP). Paper presented at the annual meeting of the American Statistical Association, 5–9 August, Atlanta.

Doyle, Pat, Elizabeth Martin, and Jeffrey Moore. 2000. *The Survey of Income and Program Participation (SIPP) Methods Panel: Improving Income Measurement.* SIPP Working Paper No. 234. Washington, D.C.: U.S. Census Bureau.

Doyle, Pat, and Jeffrey Moore. 2001. Methods Panel to Improve Income Measurement Analysis of an Experimental SIPP Instrument. In *2001 Proceedings of the American Statistical Association*. [CD-ROM]. Alexandria, Va.: American Statistical Association.

Griffiths, Julia Klein. 2001. Household Income Screening Procedures in the Survey of Income and Program Participation (SIPP) Methods Panel. In *2001 Proceedings of the American Statistical Association*. [CD-ROM]. Alexandria, Va.: American Statistical Association.

Moore, Jeffrey C. 2001. Asset Ownership Screening Procedures in the SIPP Methods Panel. Paper presented at the annual meeting of the American Statistical Association, 5–9 August, Atlanta.

Pascale, Joanne. 2001. Labor Force Participation in the Survey of Income and Program Participation (SIPP) Methods Panel. In *2001 Proceedings of the American Statistical Association*. [CD-ROM]. Alexandria, Va.: American Statistical Association.

U.S. Census Bureau. 2001. *SIPP Users' Guide*. 3rd Edition. Washington, D.C.

Disclaimer and Acknowledgments: This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. My thanks to Nancy Bates,

Jeff Moore, and Michael Morgan for assistance in preparing this report and to the entire methods panel team for their outstanding efforts to keep this project on schedule and of high quality.

Table 1: Unit Response Rates for Wave 1 by Experiment by Treatment Group

	Summer 2000		Summer 2001	
	Test	**Cont**	**Test**	**Cont**
Eligible households	1,032	988	1,041	1,120
Interviewed HHs	854	842	870	950
Noninterview (%)	17.3	14.8	16.4	15.2
People in interviewed HHs	2,170	2,122	2,266	2,519
Children (%)	21.4	23.0	24.0	23.1
Interviewed adults (%)	75.5	74.7	74.0	75.7
Noninterviewed adults/partials (%)	3.0	2.3	2.0	1.2
Proxy interviews (%)	28.9	29.4	26.3	27.8
Weighted (1000)	38,063	37,844	42,397	43,145

Table 2: Item Nonresponse Rates for Wave 1 by Experiment and by Treatment Group

Fraction of income missing	Summer 2000		Summer 2001	
	Test	**Cont**	**Test**	**Cont**
Household	.28**	.32	.22**	.28
Person labor force	.02	.02	.02**	.03
Person earnings	.23	.22	.10**	.17
Person assets before nonresponse follow up	.40**	.45	.34**	.38
Person assets after nonresponse follow up	.18**	.31	.17**	.24
Other person income	.20	.23	.18	.21

Note: There were no changes in procedures for the earnings questions administered to the treatment group in summer 2000.

* significant difference at .10 level **significant difference at .05 *** significant difference at .01

Table 3: Interviewer Preferences for Wave 1 by Experiment and by Instrument Component

Instrument Component	Summer 2000	Summer 2001
General	n/a	**Test**
Roster probes	*Control*	**Test,** slightly
Other demographic changes	**Test**	**Test**, mostly
Labor force participation	*Control*	*Control*
Unearned income	**Test**, mostly	**Test**
Asset ownership	**Test**	**Test**
Joint ownership	n/a	*Control*
Amounts - all sections	Mixed	**Test**
Health insurance	Mixed	**Test**

Table A: Labor Force Participation of Persons Age 15 and Older by Experiment and by Treatment

Weighted Distribution	Summer, 2000			Summer, 2001		
	Control (n=1,605)	**Test** (n=1,656)	**Difference**	**Control** (n=1,914)	**Test** (n=1,686)	**Difference**
Overall Differences in Labor Force Participation						
Percent of adults who are working	68.6%	66.7%	1.9	68.4%	68.9%	- 0.5
Mean number of work activities per adult[1]	.771	.760	.011	.759	.805	-0.046***
Labor Force Participation by Work Category						
Percent of adults with at least 1 wage and salary job	62.3%	58.3%	4.0	61.3%	58.1%	3.2***
Percent of adults with at least 1 business	6.7%	7.0%	-0.3	8.2%	8.7%	-0.5
Moonlighters	2.2%	0.5%	1.7	1.6%	0.6%	1.0***
Self-employed	n/a	3.3%	n/a	n/a	5.2%	n/a
Odd jobs	n/a	2.7%	n/a	n/a	2.6%	n/a
Labor Participation by number of Wage and Salary Jobs						
Percent of adults with 1 job	56.0%	52.0%	4.0%	55.2%	52.7%	2.5%**
Percent of adults with more than 1 job	6.3%	6.4%	-0.1%	6.1%	5.3%	0.8%***

* significant difference at .10 level **significant difference at .05 *** significant difference at .01

[1]Having at least one of a specific type of work, counts as a work activity. Type of work is defined as wage and salary, business ownership, other self employment, moonlighting, and other.

Table B: Mean and Median Person Earnings by Month by Type by Experiment and by Treatment

	Summer, 2000				Summer, 2001			
	Mean		Median		Mean		Median	
	Control	Test	Control	Test	Control	Test	Control	Test
Total Earnings								
Monthly total 1	$3068	$3073	$2400	$2250	$3485	$3198	$2592	$2500
Monthly total 2	$3095	$2937	$2400	$2174	$3371	$3479	$2480	$2500
Monthly total 3	$3106	$3015	$2400	$2200	$3402	$3214	$2400	$2500
Monthly total 4	$3130	$2921	$2306	$2080	$3538	$3209	$2500	$2500
Wages and Salaries								
Monthly total 1	$2939	$2932	$2384	$2280	$3349	$3110	$2500	$2500
Monthly total 2	$2968	$2782	$2400	$2192	$3215	$3246	$2400	$2511
Monthly total 3	$2978	$2877	$2400	$2200	$3285	$3132	$2322	$2505
Monthly total 4	$2981	$2756	$2240	$2100	$3334	$3129	$2400	$2500
Business Earnings								
Monthly total 1	$3920	$4651	$2320	$2353	$4217	$6156 **	$3000	$4167
Monthly total 2	$3996	$4585	$2320	$2500	$4359	$9480 *	$3000	$4000
Monthly total 3	$3919	$4692	$2400	$2800	$3773	$5814 **	$2600	$3888
Monthly total 4	$4183	$5000	$2400	$2688	$4719	$5888	$3000	$4000

* significant difference at .10 level **significant difference at .05 *** significant difference at .01