

The Survey Of Income And Program Participation (SIPP) Methods Panel--Improving Income Measurement ⁽¹⁾

Pat Doyle, Jeffrey Moore, Elizabeth Martin
U.S. Census Bureau

I. Abstract

An interdivisional group at the Census Bureau recommended new research and development to improve the instrument used in the Survey of Income and Program Participation (SIPP). As a result of that committee's work, the Census Bureau established a new project to research alternative measurement methods and test alternative approaches to asking the questions in the SIPP instrument. Under that project, alternative formulations of the SIPP questions will be tested in the cognitive laboratory as well as in the field. To rigorously test the performance of alternative approaches to the SIPP questions, staff will administer a series of field experiments to compare the performance of experimental and official SIPP survey instruments in two independent samples and then will evaluate the outcomes across the samples. One of the primary measures of success of the alternative methodologies will be the reduction of nonresponse rates to survey items that historically have been difficult to collect. A secondary measure will be a reduction in under reporting of income. This effort will be repeated three times with successively more refined experimental instruments.

II. 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. Results from the survey have far-reaching implications for national policy.

Currently, SIPP consists of 12 waves, or rounds of interviewing, with each wave administered every 4 months to a nationally representative sample of the civilian noninstitutionalized population. Interviewing for each wave is distributed over 4 successive calendar months to create a stable production workload for field staff.

The survey instrument is extremely complex, collecting information about the structure of households, economic status, sources of income, and labor force participation. The instrument consists of a core section which is repeated each wave, and "topical modules" which vary in content from wave to wave. 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 questioned (typically in Wave 1). 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, whose task was to review the SIPP core instrument to improve the instrument and, if possible, shorten it to reduce respondent burden.⁽²⁾ CIIG generated an extensive set of recommendations ranging from minor wording changes to considerable restructuring of some sections of the instrument. Recommendations were based on careful review of the instrument, on evidence about sources and magnitudes of errors in the data, and on feedback from Census Bureau field representatives about the questions that were problematic in the administration of the interview. In developing recommendations, CIIG took account of relevant methodological research and developmental work on other surveys. For example, based on research conducted for the Census Bureau's American Community Survey (Moore and Moyer, 1998), CIIG recommended that the SIPP demographic questions be restructured. Currently SIPP asks all demographic questions for one person and then turns to the next person to ask all the questions again. This is a person-based approach. The restructuring reorders the questions so that the first demographic question (or topic) is asked of all persons before moving on to the next demographic topic. This is a topic-based approach.

CIIG also recommended testing all of the proposed new approaches before implementing them in the production SIPP instrument. The need for thorough and rigorous testing led CIIG to recommend (and SIPP Executive Committee to accept) the creation of a methods panel project, separate from the production survey.

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

III. Objectives of the Methods Panel

The project's primary goals are to improve the quality of SIPP core data by improving individual items and sections of the questionnaire, by reducing nonresponse to particular survey items, and by redesigning the instrument to be more easily administered by interviewers and less burdensome for respondents. The methods panel staff set the following objectives:

- Design and evaluate alternative measurement approaches (like the topic-based approach to demographic questions) for core SIPP items in cognitive and field tests.
- Evaluate revised core instruments used in Wave 1 and Wave 2 against current SIPP instrumentation in an experimental field test to detect significant improvement in data quality or ease of administration of the survey.
- Incorporate the best performing measurement approaches in final Wave 1 and Wave 2+ core instruments and document and thoroughly test the new instrument.
- Deliver final core instruments to the Demographic Surveys Division (DSD) in time for implementation in the 2004 panel (July 2003 for Wave 1; November 2003 for Wave 2).

The topic areas which will be the focus of the research and redesign efforts include: roster questions and probes; the structure of the demographic questions; and questions on sources and amounts of income and on labor force participation, particularly among contingent and self-employed workers.

Ultimately, Census Bureau management is considering using this experimental administration of SIPP as a model for evaluating refinements of other ongoing surveys to allow testing of enhancements to the instrument while the survey is in process without disrupting the production survey.

IV. Methods Panel Study Design

We will first conduct a series of research and analytic tasks and then proceed to formal experiments with alternative questionnaires which can be evaluated to detect any improvement in underlying results or ease of collection. The research and analytic components will consist of: literature reviews (for example, on dependent interviewing and the use of event histories as a way of improving responses); cognitive testing of alternative questions in our cognitive laboratory at the Census Bureau; and analysis of existing SIPP data to ascertain the quality of the current approach to collecting information. The results of these research tasks will guide the formulation of alternative instruments which will be formally tested in the field tests of the methods panel.

The project will encompass three formal field experiments, as illustrated in the Appendix. Each field experiment will consist of a representative sample of households in six regional offices--Philadelphia, Kansas City, Seattle, Charlotte, Atlanta and Dallas--randomly assigned either to a treatment group or a control group. Each household in the treatment group will receive a modified SIPP instrument reflecting the experimental questions under review. Each household in the control group will receive the current SIPP instrument.

Experiment 1

The first field experiment will be conducted in July and August 2000 and will focus exclusively on testing the proposed

revisions to the Wave 1 instrument. The treatment group will receive a version of Wave 1 that encompasses the full set of new questions and procedures that we wish to test. The control group will receive the SIPP 2000 panel Wave 1 instrument.

Experiment 2

The second field experiment will focus on both Waves 1 and 2. A refined Wave 1 experimental instrument and the official SIPP Wave 1 instrument will be administered in June and July 2001 to new sample groups, again with approximately the same number of households receiving the test and control instruments. The refinement of Wave 1 will be to eliminate or modify question versions that proved unsuccessful in phase 1.

In October and November 2001, the same sample groups will be given a second interview. The treatment group will receive an experimental Wave 2 instrument corresponding to the Wave 1 instrument fielded 4 months previously. The primary test in Wave 2 will be the use of dependent interviewing. The control group will receive the standard version of the SIPP Wave 2 instrument.

Experiment 3

In the third phase of the experiment, staff will draw a new sample of households and administer another test of Wave 1 control and experimental instruments in July and August 2002, and a Wave 2 test in November and December. The treatment groups will receive a final experimental version of the Wave 1 and Wave 2 instruments. The control group will receive the standard versions of the Wave 1 and Wave 2 instruments.

V. Sample Design

For each experiment, we will select a sample of approximately 1,350 addresses for a test treatment and another 1,350 for a control treatment. This sample should yield approximately 1,000 households interviewed in each group (total n=2,000). Interviewing 1,000 households in each treatment would identify differences between item nonresponse rates within 3 and 8 percentage points as illustrated in Table 1. The actual detectable difference would depend on the nonresponse rate on the current SIPP instrument and the universe of households being asked the question.

To illustrate how to read Table 1, consider the question for "Interest Amount", which is administered to almost all households and which has a nonresponse rate of 10 percent in the 1992 and 1993 Panels (U.S. Census Bureau, 1998). Table 1 shows that to be statistically significant, any item that is asked of all households with a 10 percent nonresponse rate would need a nonresponse rate that is 3 percentage points lower among the test group than among the control group. The Wave 2 results are comparable to Wave 1, although the tests are slightly less sensitive. The effective sizes and detectable differences in Table 1 assume that there will be 1,000 Wave 1 households and 920 Wave 2 households interviewed in the test and control treatments, using a two-sided test with alpha equal to 10 percent. The drop to 920 households interviewed in Wave 2 takes into account expected sample attrition.

To maintain comparability between the test and control treatments, we will randomly assign sample cases so that each Field Representative's (FR) workload assignment has approximately the same number of test and control treatment households. If the FR must have either all test or all control treatment households in his or her workload, we recommend switching assignments each of the 2 months for each iteration. In other words, FRs assigned to the test treatment in the first month would be assigned to the control treatment in the second month of an iteration, and vice versa.

**Table 1. Minimum Detectable Differences in Item Nonresponse Rates
(Assumes Test and Control Treatments Interview 1,000 HHs Each in Wave 1)**

	Items Asked of All Households	Items Asked of 50% of All HHs	Items Asked of 30% of All HHs
Control Group Item Nonresponse Rate.....50%			
Smallest Detectable Difference Wave 1	4%	6%	8%
Smallest Detectable Difference Wave 2	4%	6%	8%
Control Group Item Nonresponse Rate.....30%			

Smallest Detectable Difference Wave 1	4%	5%	6%
Smallest Detectable Difference Wave 2	4%	5%	7%
Control Group Item Nonresponse Rate.....10%			
Smallest Detectable Difference Wave 1	3%	3%	4%
Smallest Detectable Difference Wave 2	3%	3%	4%

VI. Evaluation Methodology

The effects of the instrument changes on data quality will be evaluated by comparing household-level and item-level nonresponse patterns and income and program participation reporting patterns across the experimental and control groups. In addition, we will conduct cognitive research in the field, and will also assess the instruments using behavior coding and interviewer and respondent debriefings.

To incorporate refinements to each repetition of the experimental instrument and to produce a final, fully tested instrument for the 2004 SIPP panel, we must produce and evaluate methods panel results very quickly. Our goal is to produce and compare household nonresponse rates across test and control groups within one week of each experiment's close out, and item nonresponse rates within one month of close out. We will construct crude weights for both samples (basically reflecting the sampling ratio adjusted for noninterview) within two months and begin analyzing aggregate reports of income reciprocity and amounts from the weighted data. We will also conduct cognitive research directly on the methods panel samples.

Nonresponse

We will compare household noninterview rates between treatments to determine whether the new questionnaire affects household response rates. We also will compare the rates of Type Z (person refusal in an interviewed household) and partial interviews in the two treatments. Finally, among interviewed persons, we will compare rates of item nonresponse across the two treatments. We will use the figures in Table 1 to assess the significance of any differences we find.

Improved Reporting

We will compute total recipients by income source, by month and total amounts received, and compare these across the two treatment groups. We make an explicit assumption that significant increases in the aggregate statistics signal improved reporting. This assumption is justified by the fact that income sources and amounts are generally underreported, although the extent of underreporting varies by source (Moore, Stinson, and Welniak, 1999). To the extent feasible, we will attempt to assess whether the reporting improved universally across the total population or whether it was concentrated among a particular group. However, the size of the sample will severely limit how much subdividing of the population can be done.

In addition to aggregate reporting of reciprocity and amounts, we will monitor statistics on the number of transitions among reciprocity status across months, to detect any apparent change in the seam bias problem.

Other Evaluation Techniques

In addition to the cognitive work to be conducted in the laboratory, we will carry out the following research in both the test and control treatments to provide more information about data quality.

- Ask respondent debriefing questions as part of the interview to assess subjective respondent burden, privacy concerns, comprehension and recall problems, and other indicators of quality, we will compare these results between the test and control treatments, and use them to further refine the design of the new instrument.
- Use FR debriefings to assess FR preferences and detect problems in the instruments for both treatments.
- Use behavior coding of interviewer/respondent interactions and compare treatments to determine which version

causes fewer problems in administration (about 150 households per treatment).

- Use data on the time it takes to administer the instrument as a whole, and each section of it, to compare treatments using an objective measure of burden.

We will accomplish these analytic goals within the allotted time by eliminating the post-data collection processing step and by loading the output of the automated instruments directly into SAS files. The conversion to SAS has been developed for the SIPP 1996 panel and is possible to do very quickly after closeout.

VII. Research Findings to Date - Design Implications and Evaluation Plans

Asset Ownership

We are continuing to explore a number of aspects of the way in which we collect asset data. One area where we have some results is the determination of asset ownership. Data from the current (1996) SIPP panel offer support for a revised approach that reduces the burden of unnecessary questions for a substantial number of respondents without affecting the quality of the asset data we collect.

Currently, the SIPP procedures ask all respondents whether they own each of 12 asset types. In the past, interviewers have often complained that the full asset list is quite tedious and mostly unnecessary, especially in low income households, and in fact the SIPP data support this position. The data show that the overwhelming majority of respondents - over 97% - who say "no" to the most commonly-owned assets also do not own any of the less-common types. The few who do own any of the less common assets tend to realize very little income from them.

The new approach we are developing will ask all respondents about ownership of each of the six most commonly-owned assets. For respondents who report owning one or more of these assets the questions on ownership of the other types will simply continue, as in the current instrument. However, for those who report that they own none of the initial, common asset types, the remaining types will be captured by a single, catch-all "or any other financial investment" question (with detailed follow-ups only for those who respond positively).

Based on the encouragement offered by analysis of the 1996 data, we will proceed to develop and test the new approach to measuring asset ownership and will formally evaluate it in the methods panel field experiments. Evaluation of the impact of this change will focus primarily on differences (if any) between the experimental instrument and the current instrument in the rates of reported ownership of the various asset types and in total asset income reported for the wave, especially for the less-common types which may not be mentioned explicitly in the new procedure. We will also assess time of administration and subjective reactions to the instruments by interviewers and respondents to determine if this change reduces the burden on households with few or no asset holdings.

Procedures and Probes for Rostering Household Members

Revised probes designed to be used while listing the roster of household members will be subjected to cognitive and field testing. The aim is to improve population coverage in SIPP by including tenuously attached and marginal household members, who currently tend to be omitted from household rosters. Prior pilot research based on the Living Situation Survey demonstrates that marginal people will be mentioned under additional probing, and that many of those mentioned do consider themselves to be household residents, even when household respondents do not, or have insufficient information (see Sweet, 1994; Martin 1996, 1999). Additional, nonstandard probes were especially effective at eliciting mentions of minority males, who tend to be missed at relatively high rates in surveys and the census (Sweet, 1994). The probes to be tested in the methods panel are designed to stimulate mentions of commonly undercounted categories, including commuter workers and live-in employees, people who are often absent or who are mobile. Questions to determine residency status will also be included in field testing in order to screen out individuals who do not meet SIPP's criteria for residence in a sample household.

Dependent Interviewing

Dependent interviewing is used routinely in the collection of information on unit composition in surveys with repeat

visits to the same units as it reduces the size of the collection effort. This technique is used frequently in the repeated collection of occupation and industry to reduce spurious changes that might result from varying descriptions of the same task. In the case of SIPP dependent interviewing is viewed as critical for the resolution of the so-called "seam bias" problem. "Seam bias" refers to the situation common to longitudinal surveys where respondents tend to misreport the timing of their transitions, reporting that they are more likely to have occurred at the juncture of the two reference periods of consecutive rounds of interviews than at any other point during the reference period.

However, dependent interviewing is also seen as potentially problematic from the perspective of privacy policy. The nature of dependent interviewing is to recall information in the current interview which was provided in a prior interview. If the respondent changes between the two interviews, then previously reported information potentially could be revealed to someone who did not originally report it.

Mathiowetz and McGonagle (1999) prepared a review of the literature on the use of and benefits of dependent interviewing for this project. Based on that review they recommend that SIPP continue to use dependent interviewing to develop and maintain information on unit composition and other rosters, implementing a research task to assess the impact of alternative approaches on the enumeration. Mathiowetz and McGonagle also recommend continuing the use of dependent interviewing in the determination of income reciprocity but experimenting with two alternative approaches, one that reveals the prior information before asking the question and the other reveals the prior information after asking the question, if there is an inconsistency.

VII. References

Martin, E., (1996), "Household Attachment and Survey Coverage," Proceedings of the Survey Research Methods Section, American Statistical Association.

Martin, E., (1999), "Who Knows Who Lives Here? Within-Household Disagreements as a Source of Survey Coverage Error," Public Opinion Quarterly, Summer 1999.

Mathiowetz, Nancy A. and Katherine A. McGonagle (1999), "An Assessment of the Current State of Dependent Interviewing in Household Surveys." Paper prepared under Census Bureau contract #50-YABC-7-66019, (forthcoming).

Moore, Jeffrey and Laureen Moyer (1998), "Questionnaire Design Effects on Interview Outcomes." Paper presented at the Annual Meetings of the American Association for Public Opinion Research, St. Louis, MO, May 1998, and published in the Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 851-856.

Moore, Jeffrey, Linda Stinson, and Edward Welniak (1999), "Income Reporting in Surveys: Cognitive Issues and Measurement Error." In Monroe Sirken, Douglas Hermann, Susan Schechter, Norbert Schwarz, Judith Tanur, and Roger Tourangeau (eds.), Cognition and Survey Research. New York: Wiley.

Sweet, E. M., (1994), "Roster Research Results from the Living Situation Survey," Proceedings, 1994 Annual Research Conference. Bureau of the Census.

U.S. Census Bureau (1999), "SIPP Quality Profile." Washington, DC: U.S. Census Bureau.

Appendix: SIPP Methods Panel - Milestones Schedule

(1998)			(CY 1999)										(CY 2000 - part)										
FY 1999					FY 2000																		
O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S
Wave 1 instrument development/research/testing (thru 6/15/00); initial development of Wave 2																	**			Wave1 ₁			
																	^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^			(n=2000)			

(2000)		(CY 2001)										(CY 2002 - part)														
FY 2001											FY 2002															
O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S			
**											Wave1 ₂	**	Wave2 ₂	**											Wave1 ₃	**
Evaluation/refinement of Wave 1 (thru 4/15/01); "final" development of Wave 2 (thru 8/15/01)											(n=2000)			Evaluation/refinement of Wave 2 (thru 9/15/02); (cont'd eval/refinement of W1)											(n=2000)	

(2002)		(CY 2003)										(CY 2004 - part)													
FY 2003											FY 2004														
O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S		
Wave2 ₃	Evaluation/refinement of Wave 1 (thru 7/1) and Wave 2 (thru 11/1)											NEW "BIG" SIPP PANEL W1					NEW "BIG" SIPP PANEL W2								
											7/1/03 - due date for final W1 instrument					11/1/03 - due date for final W2 instrument									

TABLE KEY:

** = mid-month, 6-weeks-in-advance deadline for MP test cycle instrument delivery to TMO.

^^ = "blackout" period due to Census 2000 field activities.

(1) This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau publications. This report is released to inform interested parties of research and to encourage discussion.

(2) We thank our fellow members of CIIG, whose contributions are reflected in this paper. They include: Donna Boteler, John Bushery, Karen Bogen, Julia Klein-Griffiths, Vicki McIntire, Sean McLaren, Martin O'Connell, Joanne Pascale, and Edward Welniak.