

**THE SURVEY OF INCOME AND
PROGRAM PARTICIPATION**

**The Survey of Income and Program
Participation (SIPP) Methods Panel
Improving Income Measurement**

No. 234

Pat Doyle, Betsy Martin, and Jeff Moore
U.S. Census Bureau

November 2000

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

The Survey of Income and Program Participation (SIPP) Methods Panel Improving Income Measurement

by Pat Doyle, Betsy Martin, and Jeff Moore
November 13, 2000

Keywords: Income measurement, methods panel, experimental research

I Abstract^{1,2}

The Census Bureau recently established the Survey of Income and Program Participation (SIPP) 2000 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 income under reporting, 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 two iterations of testing and refining the Wave 1 and Wave 2+ core instruments. Each phase will culminate in a split-sample field test of about 2,000 interviewed Wave 1 households—1000 randomly assigned to each of the control treatment (receiving the standard SIPP instrument) and experimental treatment (receiving a modified instrument). We will assess our success in meeting our objectives through analysis of data obtained in the field experiments and debriefings of respondents and interviewers.

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.

¹This 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. An abbreviated version of this paper appears in the Proceedings of the American Statistical Association (Doyle, Martin and Moore, 2000).

²We thank Karen Schwager for her editorial assistance and our fellow members of CIIG, whose contributions are reflected in this paper. The CIIG members include: Elaine Hock, Julia Klein-Griffiths, and Joanne Pascale.

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 question to each person age 15 or older (or their proxy) in interviewed households.

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. 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 onto 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 research project conducted in parallel to the production SIPP and 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; questions on sources and amounts of income; and questions 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 are currently conducting a series of research and analytic tasks to be followed by a series of formal experiments with alternative questionnaires. The alternative questionnaires will be evaluated to detect any improvement in underlying results or ease of collection. The research and analytic components 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 the alternative instruments being evaluated in the field tests of the methods panel.

The project will encompass the following three formal field experiments, according to a schedule illustrated in the Appendix:

- < Experiment 1 (2000): Wave 1 control and experimental instruments administered to the first control and first treatment samples respectively
- < Experiment 2 (2001): Wave 1 control and refined experimental instruments administered to the second control and second treatment samples respectively. Each sample member is reinterviewed 4 months after the first interview using the control or experimental Wave 2 instruments
- < Experiment 3 (2002): Wave 1 control and final wave 1 experimental instruments administered to the third control and third treatment samples, respectively. Each member of the third sample is reinterviewed four months after the first interview using the control and final Wave 2 experimental instrument.

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.

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 28 to 30 percent. Table 1 shows that to be statistically significant, any item that is asked of all households with a 30 percent nonresponse rate would need a nonresponse rate that is 4 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 4 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.

V 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.

- < *Nonresponse.* We will compare household noninterview rates, Type Z (person refusal in an interviewed household) rates, and partial interview rates between treatments to determine whether the new questionnaire affects household response rates. Among interviewed people, we will compare rates of item nonresponse across the two treatments.
- < *Improved Reporting.* To the extent feasible within our sample size constraints, we will assess whether total recipients by income source by month and total amounts received increased and whether that improvement occurred universally across the total population or whether it was concentrated among a particular group. We also will search for a reduction in seam bias.
- < *Other.* We will debrief the respondents to assess difference across treatments in subjective respondent burden, privacy concerns, comprehension and recall problems, and other indicators of quality. We will also debrief FRs to assess FR preferences and detect problems in the instruments for both treatments. We will use timing data to determine if there is a reduction in the time it takes to complete the interview with the experimental instrument.

VI Research Findings to Date

The methods panel project fielded the first experiment and completed a number of research tasks related to the questions we pose in the instrument. We sponsored a review of dependent interviewing techniques. We also completed an analysis of the SIPP 1996 panel data to determine the success of the approaches to questionnaire design introduced in the 1996 panel to improve the survey data. We incorporated information from the Wave 8 1996 panel topical module on welfare participation into the core instrument.

Finally, we completed an extensive round of cognitive interviewing to test new approaches suggested by CIIG for SIPP 2004. Each of these is discussed in turn.

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 SIPP, dependent interviewing is viewed as critical for the resolution of the so-called "seam bias" problem.

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.

Analysis of 1996 SIPP

Demographic Items. In the case of housing related information (tenure and participation in public or subsidized housing) we analyzed the pattern of changes in these characteristics among units who remain at the same address over multiple waves of SIPP. We were looking to see if we could reduce the burden on respondents in the second and subsequent rounds of interviewing by skipping these questions for respondents remaining at the same address. Results showed that less than 1 percent of units change their tenure status when address remained the same so that we could safely reduce the number of times we administer the tenure question. We concluded, however, that we could not reduce the repetition of the public and subsidized housing questions in the same fashion because change in status was more prevalent among low income persons staying at the same address. (Note, however, that we will reduce the repetition of such questions anyway with the use of the income screener discussed below.)

We learned that fewer than one half of 1 percent of the children whose biological parents had the same race, reported a different race from the parents. Hence, Experiment 2 we will not ask the race question of children in the following circumstances: they reside with both biological parents and both parents are of

the same race. In those cases (and those cases only) the instrument will assign the race of the child to equal the race of the parents.

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 noted below 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 a set of most commonly-owned assets also do not own any of the less-common types. In other words, the likelihood of reporting ownership of assets from the less commonly held assets is 3 percent. Furthermore, the amount of income earned from assets held by this particular group of people was less than 8 percent of the total amount of income from assets.

Thus, we felt it was feasible to test to see if a new two-part structure could be implemented without jeopardizing the quality of the asset ownership. The new two-part series first asks if there is ownership of one of the set of common asset types. If a respondent answers yes to any in this set, then the instrument asks about the second set (of less common asset types) as in the current instrument. If a respondent answers no to all of the first set, then the instrument poses a catchall question as to whether any other income-producing assets were owned. If a respondent answers yes to the catchall question, then the instrument asks about the less common asset types individually, as in the current instrument.

Another interesting fact we learned from the analysis of 1996 panel data was that most of the nonresponse on asset income consisted of “don’t know” rather than “refused.” Overall, respondents indicated that they did not know the amount twice as often as they refused to provide an amount. Thus, we believe we can improve the collection of asset income by expanding our use of nonresponse follow-up questions, as has been recommended by the Census Advisory Committees.³

Further, we examined the nonresponse follow-up questions already included in SIPP (there are six asset types for which a nonresponse follow-up question is included in the current instrument) and found them to be successful in determining an amount for two-thirds of the cases of initial nonresponse. The format of those follow-up questions was to ask if the amounts fell within certain brackets. For those followup questions we are updating the brackets used for the methods panel based on the distribution of asset income from the 1996 panel.

³Refer to Juster and Smith (1997) for a discussion of one form of nonresponse followup suggested by the advisory committee.

Seam Bias. We have explored and continue to explore the success of the 1996 panel in reducing the seam bias. As illustrated by the graphs in Figure 1 we were not entirely successful in reducing the occurrence of the seam bias. For some income sources (like AFDC and food stamps), the seam bias remained at the same levels as the 1993 panel. For some income sources (like SSI and Pensions) proportionately more transitions occurred at the seam in the 1996 panel as in the 1993 panel. Finally, for the remaining income sources (like State Unemployment Compensation and WIC) the outcomes were reversed.

In the 1996 panel the attempts to improve the seam bias were focused on questions that will have their biggest impact in post collection processing. Since the outcome in Figure 1 were based on raw data, we decided to see how much the estimates could be improved by editing the information during post collection processing. We learned that if we assumed that reports of reciprocity for the interview month of the prior wave were correct (instead of the no reports for the same month from the next wave), then we could shift about half of the transitions away from the seam. Of course the shift is only by one month so that the revised graphs still look like the graphs in Figure 1 but they have plateaus rather than peaks (see Figure 2).

We are currently examining the other questions in the 1996 panel (the so-called previous wave questions) that condition the question on reciprocity in wave $n+1$ on reciprocity in wave n . The expectation in designing the 1996 panel was that the results of the previous wave questions would yield transition dates distributed more evenly across the between-wave and within-wave transition points. Unfortunately, we are finding that most of the information gathered in wave $n+1$ either: indicated no transition occurred, negated the reciprocity reported in the prior Wave, left the transition at the seam, or was not useful (such as reporting the last date of receipt to be years before the survey started). Clearly we need to improve the methods for using dependent interviewing in this context.

Incorporating Wave 8 Results

The Office of Management and budget (OMB) interagency committee on SIPP formed a subcommittee to evaluate how SIPP could capture program participation and benefits in a post-welfare reform environment. That committee developed a topical module for Wave 8 of the 1996 SIPP panel that tested the following: an income screener limiting questions to persons in low-income households, a series of questions to capture participation in a variety of types of programs not explicitly captured in the SIPP core, and measures to permit valuation of benefits received under the program. That module was also designed to give us information on the extent of the use of electronic methods of providing benefits (like debit cards) instead of checks.

A series of papers presented at the American Statistical Association meetings in August 1999 (Griffiths et al., 1999, Nelson and Doyle, 1999, and Ollerich and Hauan, 1999) incorporated a description and evaluation of the attempt to capture program participation and benefits in a post welfare reform environment and recommendations for changing SIPP to better capture program participation. Of note are the findings from Griffiths et al.

Successful Income Screener. Only 87 out of 51,549 respondents received need-tested benefits and reported income in excess of the cut off. Note these results should improve with a refined screener that explicitly uses the correct cut off. The cut off used in Wave 8 was intended to be twice the poverty level but it was not computed accurately for large households (it was lower than twice poverty).

Good Response Rates for New Programs. The questions on programs not previously identified in SIPP appear to be of good quality. The authors demonstrate that except for the questions on community service the nonresponse rates are below 2 percent.

Sizeable Frequency of Receipt for the New Programs. The authors demonstrate that while receipt of any of the new programs is rare relative to the total population, we captured a substantial number of new participants. For example, more than 25 percent of the reports of receipt of new benefits are from people not reporting receipt of need-tested benefits in the SIPP core. Further, the number of participants in nearly all of the new programs is sufficient to warrant identification of these income sources on the public use data files (which cannot be said for all of the income sources currently in the SIPP core.)

Consistent Follow-up. Analysis of the follow-up questions on source, type, and amount of benefits received shows that most of the “yes” reports to reciprocity questions are indeed valid reports. Only job subsidy, job search, and food assistance (other than food assistance programs in the core) show 10 percent or more recipients failed to provide any information in the follow-up questions. Food assistance reciprocity was substantial in size (over 400 respondents reporting reciprocity) so that even ignoring those without valid follow-up questions, the additional benefits reported are significant.

Primarily Welfare. Finally, the information from Wave 8 suggests that most of the program benefits under the new programs did constitute welfare benefits from a government (as opposed to charitable or private) source.

Results of Cognitive Testing

Our cognitive testing covered the following instrument changes: new roster probes, new citizenship questions, re sequencing of labor force questions, flexibility in reporting amounts of income received, a new income screener, revised flow of the asset reciprocity questions, new nonresponse follow-up questions, a new approach to collecting months of participation, and revised wording to the school lunch program questions. We used a paper instrument to conduct the research with 26 individuals participating. The results follow.

Demographic Characteristics. We cognitively tested new rostering questions to see if they better communicate household membership rules (for example, college students for whom a room is maintained at the sample address should be included on the roster), and improve the identification of tenuously-attached household members and other people who are likely to be undercounted. We also cognitively tested revised questions which identify the person who owns or rents the home, proposed new questions

that establish citizenship, and a new two-part approach the educational attainment question. Finally, we tested education questions on enrollment and financial assistance.

Overall, the results were good. The roster probes worked well to help respondents think about tenuously attached household members. Respondents did not seem to have any problem identifying the individual(s) who owns or rents the home. The screening question for U.S. born citizens worked well to skip the majority of household members with all native born citizens to the next section. The two-part approach to educational attainment also worked well, with more improvement expected when the skips are automated. We did learn that the respondents did not seem to have a clear understanding of the response categories used for types of financial assistance and did not seem to know the differences between them. However, we did not have sufficient observations responding to these questions to guide us in making improvements. Clearly, more research is needed.

Labor Force. We tested a new approach to determining the type of employment related activities with the goal toward distinguishing business ownership, other types of self employment, wage and salaried jobs, and other less regular forms of employment. Instead of asking first about wage and salaried jobs that are most prevalent (as is currently done in SIPP), we first asked at the household level about family owned businesses and who worked for those businesses. Then we moved to the person level to identify other forms of self employment. The questions about wage and salaried employment were then administered to each adult followed by items on odd jobs. The approach worked well in cognitive testing and we are proceeding to test it in replicate 1 to determine its impact on the reporting of labor force activity.

Earnings. We attempted to allow the respondent flexibility in reporting amounts, allowing them to pick an “accounting period” that they found most comfortable. However, the version of the question we used did not entirely succeed. Instead of answering a question on how they could provide the amounts, they simply provided the amounts. Clearly, we need to do some more work on this approach. There also appeared to be some problem deciding whether/how to report money received rather than money earned in a month. SIPP’s focus is on the former.

We tested a question designed to elicit business income other than the monthly earnings already reported. This question posed comprehension problems for most respondents, but no real theme about the source of the problem was common across respondents. One respondent said she thought this was asking about “under the table” income, another said “other than what [income]?” another said it meant income not yet reported to the interviewer, and other respondents were simply confused by the question.

On a brighter note, we did observe that respondents are using estimation strategies for computing monthly income that make sense. Once respondents understood the question on amounts (which asked for amounts received thus far this month followed by each month in the reference period, working backwards) providing the amounts did not seem to be a problem. Most respondents’ reports of amounts were based on payments for discrete one-time contracted jobs. However, one respondent’s report of amounts was simply the profit from his business.

Assets. We cognitively tested some new approaches to determining asset ownership. First we tested to see if asking about IRAs and other such accounts in advance of questions on other assets would alleviate confusion as to how those assets should be counted. SIPP is interested in income earned from assets currently held but retirement accounts do not actually yield any disposable income while they are still being built up. Hence, when SIPP asks about stocks, we do not want the respondent to report stocks held in their retirement accounts. There were no real problems with this approach but we were not 100% successful in keeping the retirement accounts out of the subsequent reciprocity categories. The field test will tell us whether the frequency of confusion is reduced with the new approach.

We also tested the new two-part approach to determining asset ownership discussed earlier and some revisions to the joint versus sole ownership. These worked well with some minor exceptions and we will proceed to field test the approach to see if it improves the ease with which the interview is conducting without any impact on the overall reporting of asset reciprocity. Among the minor issues we found some confusion over our choice of words (e.g., one respondent misinterpreted the word “alone,” thinking we meant “a loan”) and we realized that we needed to ask savings bonds as part of the first set of assets (even though it is not one of the more commonly held) so that we reduce the confusion between those and other types of bonds.

General Income other than Earnings and Asset Income. We tested the series of questions from the Wave 9 of SIPP 1996 Panel on need-tested assistance programs. However, we changed those questions subsequent to the cognitive interviewing. Hence, this text focuses on other aspects of the cognitive work on general income. One focus of the test was to see if we could administer a question on receipt of disability income to all adult respondents rather than just those indicating some form of disability in earlier sections of the instrument. We suspected the disability reciprocity item over relies on the prior “work limits” question to identify potential recipients of disability income, and thus may contribute to the underreporting of disability income such as workers’ compensation. We found no problems with this expanded universe and thus plan to test it further in Experiment 1.

We tested some modifications to the approach to determining the point(s) in time when income or benefits were received, both in the current reference period and in the past. We asked if the respondent received anything yet this month (i.e., the interview month) and if so how much. Then we asked if they received it in any of the other months of the reference period. In this case we listed the months in chronological order (as that made more sense in the context of the questions) rather than asking about the most recent month first. With some minor exceptions these worked well and we will field test that approach in the first experiment. We have some issues about questions on child support pass through payments that require some further research to be conducted as part of Experiment 2.

Health Insurance. We cognitively tested respondents’ ability to identify coverage under Medicaid or government-sponsored medical assistance programs, to provide the name of the program so that we can

use it in subsequent questions probing for details on the program benefits, and to report the period during which they were covered within the reference period of the survey.

Generally, respondents seemed to have no trouble identifying themselves or their children as enrolled in Medicaid in response to the two basic Medicaid/government assistance program questions. In follow-up questions that determined the program name, results were mixed. In some cases respondents recognized one or more program names from the list provided; in other cases respondents offered their own program name which was not on the list. In three cases respondents offered multiple names for the programs they or their children were enrolled in. Almost all respondents, however, identified the program as Medicaid or the generic "Medical Assistance." This seems to indicate that for the most part, respondents refer to these programs as Medicaid and that providing a list may or may not aid in reporting.

Most respondents had no trouble answering the coverage question with certainty so we will retain the question. However, we will determine whether this question is interpreted as measuring coverage *at any time* during the month or *continuous* coverage throughout the month. If the interpretation is coverage at any time during the month, consider modifying the wording to make this more clear to respondents.

VII Experimental Instrument Design

Figure 3 summarizes the resulting changes to the Wave 1 instrument fielded in August and September 2000 (Experiment 1). These changes are listed by instrument section which are listed in the order in which these sections are administered in the field. With some exception these changes represent no new content. Further, with some additional exceptions, these changes do not yield substantial changes to the instrument output which is the input to the post collection processing. The exceptions are noted in Figure 3.

Procedures and Probes for Rostering Household Members

We introduced revised probes designed to be used while listing the roster of household members. 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 mention 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.

Demographic Characteristics

We restructured the questions to implement the topic-based approach to collecting demographic information. We also introduced a few questions designed to determine citizenship based on place of birth.

Labor Force Characteristics

We are experimenting with a different structure for the labor force participation question to try to improve our ability to capture marginal attachments to the labor force and unusual patterns of work and earnings receipt. The changes are concentrated in the treatment of households where some individuals are self-employed. We begin with questions at the household level to determine what, if any, family-owned businesses exist and the characteristics of such businesses. We ask expanded self-employment questions as a vehicle to capture often missed irregular/odd job type work and to eliminate confusion about whether this work constitutes a business.

General Income Reciprocity

Our approach to collecting participation in need-tested programs differs between the experimental and control instruments in three ways. First, we introduced a screener question to the experimental instrument to alter the flow of the instrument for persons in households with relatively high incomes (over twice the poverty line). For such households we will administer general questions about receipt of need-tested questions and, if they indicate no participation in these programs in general, we will not ask them further detailed questions. Second, we expanded the content of the experimental instrument to explicitly reference all of the types of need-tested programs known to exist in the post-welfare reform environment⁴ and to collect information on the value of those programs when the benefits are cash or cash-like⁵. Third, we integrated the questions from the “Programs” section of the instrument with the General Income Reciprocity section of the instrument. The use of a separate section for a few programs (housing and some but not all nutrition programs) in the control instrument design was a hold over from the paper questionnaire where it was easier to restrict some questions to the household reference person if they were grouped together in one module. With automated instruments, of course, that is no longer an issue.

⁴Welfare reform refers to the legislation enacted in 1996 Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) which replaced the federal-state matching program (Aid to Families with Dependent Children (AFDC)) with a federal-to-state block grant program (Temporary Assistance to Needy Families (TANF)). The major impact of this legislation from the data collection point of view was that there was no longer a national program providing cash benefits. Instead, there is a series of state and local programs providing benefits of all types and these programs are evolving over time.

⁵Reciprocity questions for some of these additional programs exist in the control instrument as well since they were recently added to the production SIPP instrument. However, the questions differ in format, sequence, and scope from those incorporated into the experimental instrument.

Asset Ownership

In the case of asset ownership, we reoriented the experimental instrument in two ways. First, we shifted questions about individual retirement and related accounts so that they would occur before questions on the ownership of stocks and other investments. The expectation is that this reorganization will help clarify that investments held in retirement and related counts are not to be included in the set of questions on income-producing assets. Second, we broke the list of questions about ownership of income-producing assets into two sets. The first set will be administered to everyone and the second set will only be administered to a subset of respondents. Participants who indicate no asset ownership based on the questions in the first set will be administered one catchall question regarding the asset types in the second set. If they say no to the general catchall question, they will not be subjected to detailed questions on the assets in the second set.

Evaluation of the impact of this change will focus primarily on differences (if any) between the experimental instrument and the control 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.

Income

In the sections that capture income amounts received we are striving to make it easier for the respondent to report actual amounts received. In the case of earnings, our attempts to do so have not yet yielded a series of questions that produces the same output as we produce in the traditional core instrument so we have postponed work on that until Experiment 2. In the case of asset amounts, we are allowing respondents more flexibility in reporting the amount questions so that they can report quarterly amounts, annual amounts, or other variations on that theme. We are also revising the flow of the questions addressing jointly held assets and expanding the nonresponse follow-up questions.

Health Insurance

In the health insurance section, we are adding questions to the content designed to capture the new Children's Health Insurance Program. We also modified the series of questions on Medicaid participation to make it easier for the respondents who seem to recognize the concept of Medicaid or medical assistance but may not be fully cognizant of the actual program names.

VIII Conclusion

The methods panel project is well underway and providing interesting results about the good quality of the SIPP program. We are finding ways to streamline the instrument to help reduce respondent burden and we are identifying areas where improvements can be made in the methods with which the information is collected. To date we learned the following:

- < the seam bias problem has not yet been solved and thus needs further research
- < use of nonresponse followup improves reporting of income amounts
- < nonreponse to asset income questions is primarily the result of lack of knowledge of the amount rather than refusal to reveal the amount, suggesting we will see significant improvement in the overall results with increased use of nonresponse follow-up questions
- < there exists a common set of asset types that can be used to reduce the burden of the determination of asset ownership by type,
- < an income screener can be successfully used to reduce the number of respondents subjected to questions about need-tested programs.

The project fielded an experimental and control instrument side-by-side in August and September 2000. The results will provide an excellent opportunity to examine the success or failure of our new methods of data collection reflecting alternative wording and approaches to items in virtually all sections of the SIPP instrument.

References

- Doyle, Pat, Betsy Martin, and Jeff Moore (2000) "Methods Pane to Improving Income Measurement in the Survey of Income and Program Participation (SIPP)." Proceedings of the American Statistical Association. Alexandria VA: American Statistical Association.
- Griffiths, Julia Klein, Jeffrey C. Moore, and Karen Bogen (1999) "Development and Evaluation of the New SIPP Wave 8 Welfare Reform Questions " Proceedings of the American Statistical Association. Alexandria VA: American Statistical Association.
- Juster, F. Thomas and James P. Smith (1997) "Improving the Quality of Economic Data: Lessons from the HRS and AHEAD." Journal of the American Statistical Association Volume 92, No. 440.
- 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 Lauren 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.
- Nelson, Charles T. and Patricia Doyle (1999), "Recommendations for Measuring Income and Program Participation in the Post Welfare Reform Era " Proceedings of the American Statistical Association. Alexandria VA: American Statistical Association.
- Oellerich, Donald T. and Susan Hauan (1999) "The Role of the Survey of Income and Program Participation Wave 8 Topical Module in Agency Evaluation of Welfare Reform" Proceedings of the American Statistical Association. Alexandria VA: American Statistical Association.
- Sweet, E. M., (1994), "Roster Research Results from the Living Situation Survey," Proceedings, 1994 Annual Research Conference. U.S. Census Bureau.
- U.S. Census Bureau (1999), "SIPP Quality Profile." Washington, DC: U.S. Census Bureau.

**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 Households	Items Asked of 30% of All Households
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%

Figure 1

1996 and 1993 Panel Reciprocity Transitions

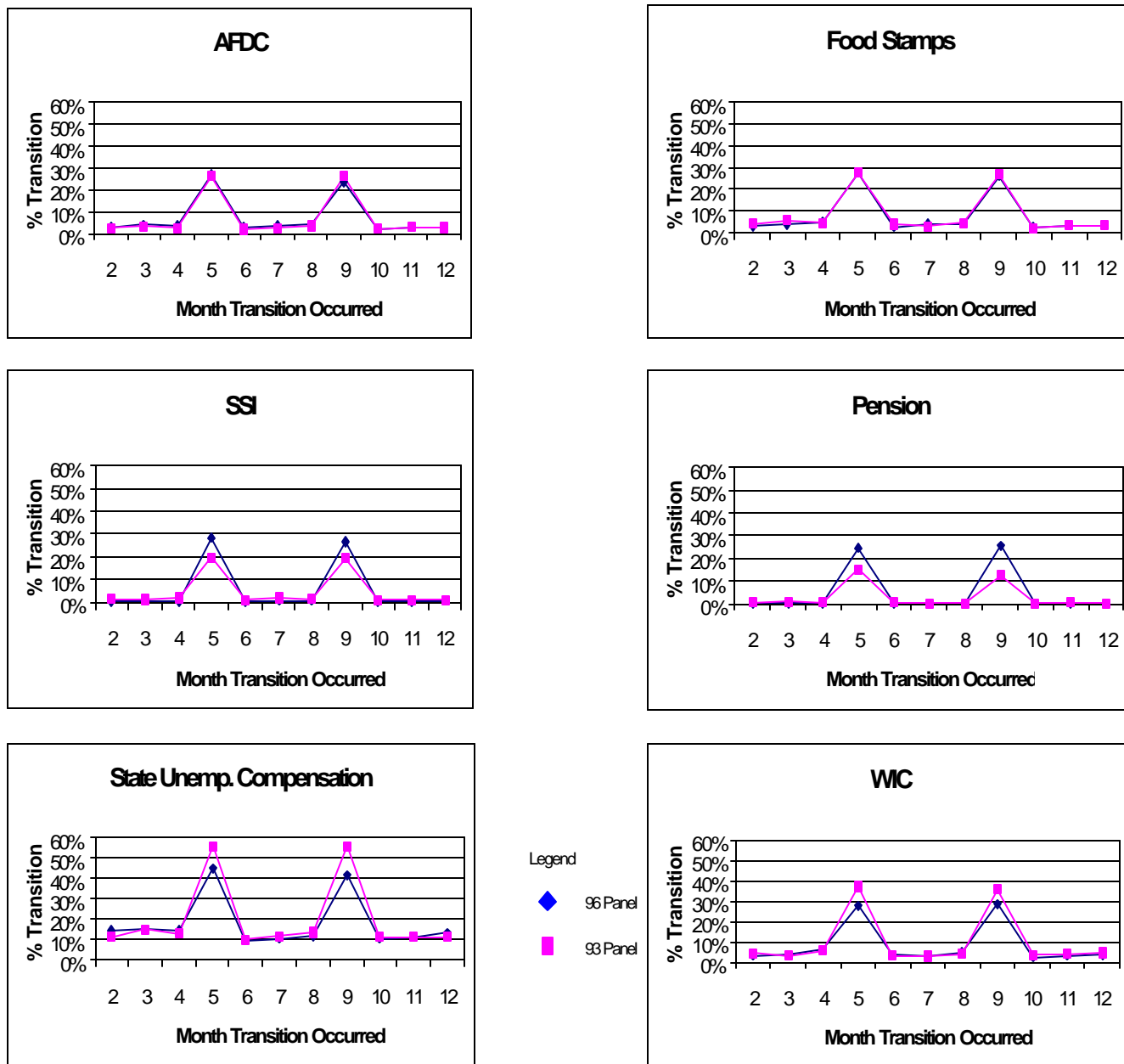


Figure 2
1996 SIPP Transitions in WIC Reciprocity, Original and Edited

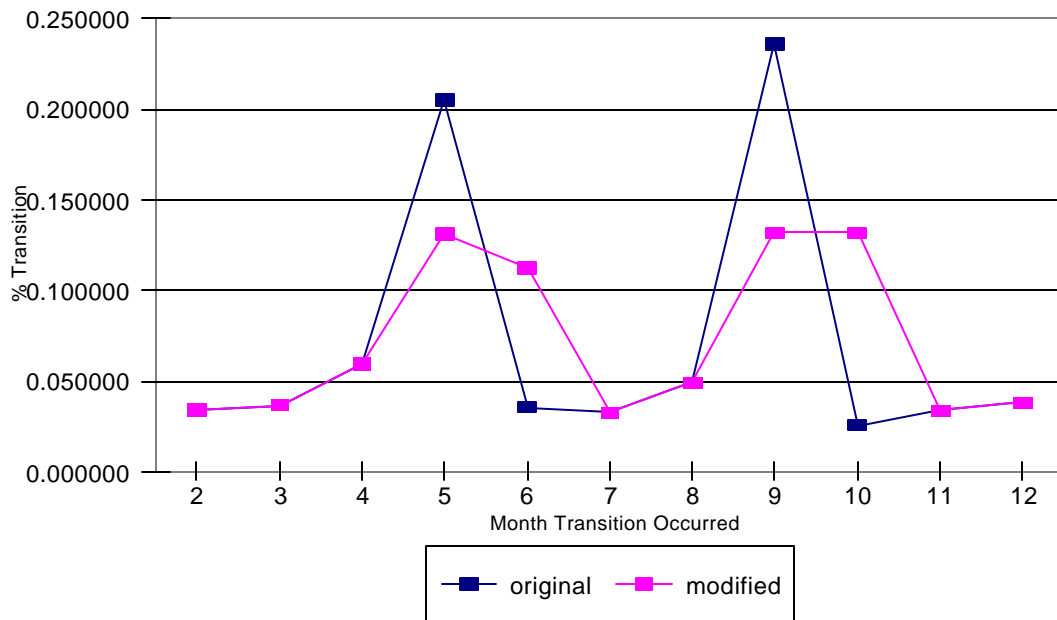


Figure 3
Summary of SIPP Wave 1 Instrument Changes to be Investigated
in the SIPP Methods Panel Research Project

Demographics	<p>Use a topic-based format (e.g., "What is NAME1's date of birth?" "How about NAME2?" "And NAME3...?" etc.).</p> <p>In rostering, capture volunteered information about relationships in order to eliminate the need for some questions.</p> <p>Identify all eligible owners/renters, select a reference person based on the owner/renter status.</p> <p>Use new approach to capture people whose circumstances often result in their being missed or improperly enumerated.</p> <p>Conduct research on improved methods for capturing educational attainment which reduce screen clutter, and reduce confusion.</p> <p>Add new questions to determine citizenship status.</p>
Labor Force	<p>Establish family-owned businesses at the household level rather than the person level.</p> <p>Eliminate duplicate questions concerning the general characteristics (e.g., the date the business was started) of the same family-owned business.</p> <p>Add questions about "self employment" in order to capture often-missed irregular/odd job type work.</p> <p>Modify and add questions in order to clarify the kinds of work categories respondents fall into.</p>
General Income Receipt	<p>Use screening procedures to avoid asking the full array of detailed, means-tested program questions in obviously ineligible households.</p> <p>Add questions to capture new forms of means-tested program participation resulting from welfare reform.</p> <p>Eliminate household-level "Programs" as a separate instrument section, and incorporate most of those questions (with appropriate skip logic) into the General Income section.</p> <p>Eliminate the required tedious repetition of the reference period by enclosing the reference period text in parentheses for most questions, signaling to FRs its optional use as needed.</p>
Asset Ownership	<p>Minimize questions about relatively obscure and uncommon asset types for respondents who do not own any of the most common types of assets by employing an assets screener question.</p>

Earnings	Develop alternative instrument paths which allow more flexibility in the reporting of earnings amounts, according to respondents' preferences, but which still conform to SIPP's need for monthly data (to be incorporated in a later test). Add questions to capture income from irregular/odd job labor.
General Income Amounts	Consolidate and avoid repetition of screens which verify reports of unusually large amounts. Update program labels, including state or local program names. Add questions to collect amounts from new, welfare-reform-related programs (if possible).
Asset Amounts	Develop alternative instrument paths which allow more flexibility in the reporting of asset income amounts, according to respondents' preferences, but which still conform to SIPP's data needs. Capture more detailed information about joint asset ownership arrangements, especially for joint with other-than-spouse ownership. Expand and clarify the collection of asset amounts for different combinations of asset owners (e.g., joint with spouse and child, joint with spouse, joint with someone outside the household) for better navigation through the section, and clearer delineation of income from different joint ownership arrangements. Explore new procedures for more efficient capture of jointly held asset income. Expand the use of amount range reporting options as a fallback option in the event of nonresponse.
Health Insurance	Add and/or modify items to more accurately measure participation in government health insurance plans, especially the new Children's Health Insurance Plan (CHIP).
Program Questions	(see above; now mostly incorporated into General Income): Clarify the "school meals" questions.
Throughout the Instrument	To assist in controlling the "seam bias," capture month 5 (interview month) educational attainment, labor force and program participation, asset ownership, and income amounts.

Appendix: SIPP Methods Panel – Milestones Schedule

(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								
** Evaluation/refinement of Wave 1 (thru 4/15/01); "final" development of Wave 2 (thru 8/15/01)						Wave1 ₂ (n=2000)		**		Wave2 ₂		** Evaluation/refinement of Wave 2 (thru 9/15/02); (cont'd eval/refinement of W1)						Wave1 ₃ (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

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

"blackout" period due to Census 2000 field activities.	
Subscript 1,2,3	Experiment 1,2,3