THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

An Overview of the Survey of Income and Program Participation, Update 1

No. 1

Dawn Nelson, David B. McMillen, and Daniel Kasprzyk

U.S. Census Bureau

December 1985

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

THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

AN OVERVIEW OF SURVEY OF INCOME AND PROGRAM PARTICIPATION UPDATE 1

No. 8401 | SIPP Working Paper # 1

Dawn Nelson, David B. McMillen, and Daniel Kasprzyk

December 1985

U.S. Department of Commerce BUREAU OF THE CENSUS

Acknowledgments

This paper was prepared by Dawn Nelson, Demographic Surveys Division, David B. McMillen, and Daniel Kasprzyk, Population Division. We would like to thank our colleagues in Population Division, Demographic Surveys Division, and Statistical Methods Division for their helpful comments on various drafts of this paper.

Suggested Citation

Nelson, Dawn, et al, D.B. McMillen, and D. Kasprzyk, "An Overview of the Survey of Income and Program Participation, Update 1," SIPP Working Paper Series No. 8401, Washington, D.C.: U.S. Bureau of the Census, 1985.

Preface

This paper is intended to replace SIPP Working Paper No. 8401--"An Overview of the Survey of Income and Program Participation," and should provide a general introduction to the survey. It incorporates information on several changes that were made in the SIPP program after the original paper was issued in June 1984. We assume that other SIPP program changes will occur after this paper is released and hope to issue additional "Updates" as frequently as necessary to describe the current status of SIPP.

Working Paper No. 8401 was written mainly by Dawn Nelson, David McMillen, and Daniel Kasprzyk. The revisions to this paper, which resulted in Update 1, were made by Dawn Nelson. Numerous other Census Bureau staff members and SIPP data users contributed advice and assistance in making these revisions.

TABLE OF CONTENTS

Page

Preface v

Text

Introduction 1

Design Features 3

Survey Content 8

Operational Procedures 11

SIPP Publications 14

SIPP Public-Use Microdata Files 15

Text Figures

1. Overlapping Panels 4

2. Relationship Between SIPP Interview Months and Reference Periods: 1985 Panel 7

Appendices

A: SIPP Panel Sample Selection Procedures 19

B: Differences Between SIPP Panels. 21

C: Explanation of Income Concepts and Noncash Benefits. 27

D: Links to Administrative Records and the Validation of Social Security Numbers 29

E: The SIPP Identification Numbering System 31

F: SIPP Edits 35

G: Overview of Cross-Sectional Weighting for SIPP 37

H: Cross-Sectional Imputation System for Item Nonresponse 39

I: Data Examples. 41

J: Schedule for Releasing SIPP Quarterly Reports and Core Public-Use Microdata Files. 45

K: SIPP Quarterly Report Tables 47

Appendix Figures

B-1. Survey of Income and Program Participation Topical Module Schedule. 24

C-1. Income Sources Included in Monthly Cash Income 28

E-1. Illustration of the SIPP Identification Numbering System 34

H-1. Cross-Sectional Imputation System Flowchart. 40

References. 49

Introduction

In October 1983, the Bureau of the Census conducted the first interviews in the Survey of Income and Program Participation (SIPP). The SIPP is a nationwide survey designed to provide comprehensive information on the economic situation of households and persons in the United States. This survey is the first to collect information on cash and noncash income, eligibility and participation in various government transfer programs, labor force status, assets and liabilities, and many other topics on a regular basis to learn how changes in people's lives affect their economic well-being. The data collected in SIPP will be especially useful for studying Federal government programs including cost and effectiveness, and the effect that proposed changes in program regulations and benefit levels will have on recipients. For example, statistics not previously available such as multiple recipiency of benefits from all major transfer programs will be produced. And, questions such as "how many persons and families eligible for benefits are not receiving them?" can be addressed with these data. Studies of other important national issues, such as tax reform, Social Security program costs, and national health insurance also can be enhanced by the information from this survey.

Demands for this type of information surfaced in the late 1960s and early 1970s when many government programs were expanded and reorganized. To carry out this work, policy researchers developed microsimulation models to measure the distributional impacts of various alternative proposals. For example, the effects of changes in eligibility rules or benefit levels on different demographic groups could be determined or the results of alternative taxing arrangements could be observed. These models, however, required detailed data to ensure the reliability of the model estimates; e.g., data on monthly income, sources of noncash income, changes in household composition, assets and liabilities, and so forth. Since no single survey collected such information, the analysts had to make many assumptions and imputations based on data from surveys designed for other purposes. The most widely used source of income data was the March Income Supplement of the Current Population Survey (CPS). These data, however, suffered from underreporting and misclassification by respondents. The only feasible way of obtaining improved measures of cash and noncash income and previously unavailable program-related data was to start a new survey specifically for this purpose.

SIPP was allotted in the budget of the Census Bureau. Work began almost immediately in preparation for starting the survey in October 1983. The following sections of this paper describe SIPP's design, content, procedures, and publicly available data.

Design Features

The primary goals in designing SIPP were to improve reporting of income and other program-related data and to do it in a way that would allow the analysis of changes over time at a microlevel. The design also had to accommodate the collection of a large quantity of information in a flexible manner that allowed some information to be collected more frequently than other information. These goals were met principally by using a survey design in which the same people are interviewed more than once. Persons at households selected for a sample panel are interviewed about their income and other topics once every 4 months for approximately 2 1/2 years. Sample persons are interviewed at new addresses if they move, and any other persons that they move in with, or vice versa, are also interviewed. In this way, a highly detailed record is built up over time for each person and household in a sample panel. This design minimizes the need for sample persons to recall most of the information for longer than a few months and reduces the number of questions asked in one interview.

To further enhance the estimates of change, particularly year-to-year change, a new sample panel is introduced every year instead of at the conclusion of a panel. Consequently, two or sometimes three panels are in the field concurrently, as is illustrated in Figure 1. Since portions of the sample are the same from one year to the next, year-to-year change estimates can be based in part on a direct comparison across 2 years for the same individuals . This design gives a more precise estimate of change than a design involving interviews 1 year apart with two different groups of individuals in which greater sampling variability obscures the actual change. This overlapping panel design also allows cross-sectional estimates to be produced from a larger, combined sample that is about double in size when 2 panels overlap and triple with 3 overlapping panels.

Consequently, the Department of Health, Education, and Welfare initiated the Income Survey Development Program (ISDP) in 1975 to research the best way to design a survey to meet these data requirements. Between 1975 and 1981, extensive new procedures were developed and tested for collecting income and related socioeconomic data longitudinally on a subannual basis. Much of the work centered around four experimental field tests that were conducted in collaboration with the Census Bureau to examine different concepts, procedures, questionnaires, and recall periods.

Two of the tests were restricted to a small number of geographic sites; the other two were nationwide. In the first nationwide test, the 1978 Research Panel, approximately 2,000 households were interviewed. Because of the relatively small number of interviews, controlled experimental comparisons of alternatives were not possible; however, the panel did demonstrate that many of the new ideas and methods were feasible. It also laid a foundation for the largest and most complex test, the 1979 Research Panel. This panel consisted of a nationally representative sample of 7,500 interviewed households and provided a vehicle for feasibility tests and controlled experiments of alternative design features.

A more complete description of the ISDP and the initial findings from the 1979 Research Panel are contained in Yeas and Lininger (1981). Additional information on ISDP, including a bibliography of ISDP materials, can be found in The Technical, Conceptual, and Administrative Lessons of the Income Survey Development Program (David, 1983). To request tapes and documentation describing the history of the 1979 Research Panel, sample design, survey content, estimation procedures, data collection and processing procedures, write to: Department of Commerce, National Technical Information Service, 5385 Port Royal Road, Springfield, VA 22161 or call (703) 487-4807.

Based on experience obtained in the ISDP, planning began for implementation of the new survey now known as SIPP. In the fall of 1981, however, virtually all funding for research and planning leading to SIPP was deleted from the budget of the Social Security Administration (the office in HEW, now HHS, responsible for SIPP). The loss of funding for fiscal year 1982 brought all work on the new survey to a halt. In fiscal year 1983, however, money for initiation of

The first SIPP panel, designated as the 1984 Panel but implemented in October 1983, started with approximately 20,000 interviewed households. The second panel, i.e., the 1985 Panel, began in February 1985 with around 14,000 interviewed households. Panels about this size are expected to be fielded every February. The sample size changes in each wave of a panel due to losses through attrition and gains from following movers to new households. See Appendix A for more information on the sampling procedures, and Appendix B for a description of some major differences between SIPP panels.

The reference period for the primary survey items is the 4 months preceding the interview; for example, in February, the reference period is the preceding October through January. When the household is interviewed again in June, the reference period is February through May. To create manageable interviewing and processing work loads each month instead of one large work load every 4 months, the sample households within a given panel are divided into four subsamples of nearly equal size. These subsamples are called rotation groups, and one rotation group or one-fourth of the sample is interviewed each month. Thus, it takes 4 consecutive months to interview the entire sample. This 4-month period of interviewing is called a "wave." However, each panel has one short wave lasting 3 months in which only three-fourths of the sample is administered the same questionnaire. Therefore, households in one rotation group are interviewed only seven times whereas households in the other three groups are interviewed eight times during the 2 1/2 years. (The 1984 panel will last almost 3 years, resulting in eight or nine interviews altogether.) The short wave is necessary to get the interviewing on a cycle that allows the questions concerning taxes and annual income to be asked in May through August each year when W-2 and IRS forms are likely to be available to help respondents supply the information. (See the following section --"Survey Content.")

The following is an illustration of the relationship between waves, rotation groups, interview months, and reference periods in the 1985 panel. The basic relationships are the same in subsequent panels although the short wave may occur at another time. Looking at Wave 1 in Figure 2, persons interviewed in February report data for the period October through January; in March another rotation group reports for November through February, and so forth for each of the 4

Figure 1.
Overlapping Panels

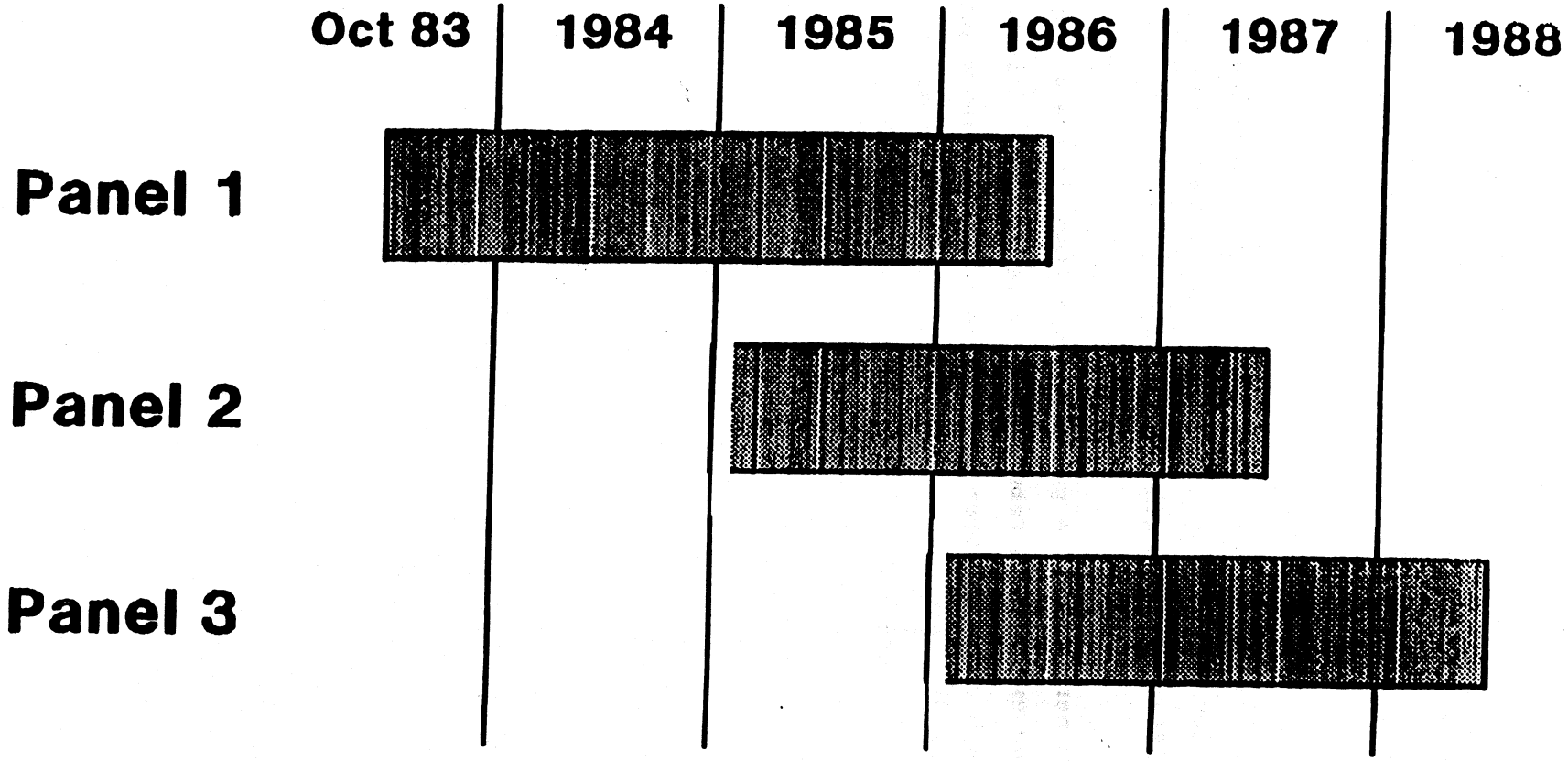


Figure 2. Relationship Between SIPP Interview Months and Reference Periods: 1985 Panel

Wave	Rotation Group	Interview Month	Reference Periods: 1st Qtr Data	2nd Qtr Data	3rd Qtr Data	4th Qtr Data	
	2	FEB	JAN				
1	3	MAR	JAN FEB				
	4	APR	JAN FEB MAR				
	1	MAY	JAN FEB MAR	APR			
	2	JUN	FEB MAR	APR MAY			
2	3	JUL	MAR	APR MAY JUN			
	4	AUG		APR MAY JUN	JUL		
	1	SEP		MAY JUN	JUL AUG		
	2	OCT		JUN	JUL AUG SEP		
3	3	NOV			JUL AUG SEP	OCT	
	4	DEC			AUG SEP	OCT NOV	
	1	JAN			SEP	OCT NOV DEC	
	2	FEB				OCT NOV DEC	JAN
4	3	MAR				NOV DEC	JAN FEB
	4	APR				DEC	JAN FEB MAR
5							
6			Continues in the pattern established above.				
7							
8							

rotation groups. Notice that each rotation group within a wave uses a different reference period, namely, the four months preceding the interview month. As a result, data are available for 7 months at the conclusion of Wave 1 interviewing although each month is not represented by the full sample. For example, the October and April data will only be available for one rotation group (rotation groups 2 and 1 respectively); the November and March data for two rotation groups (rotation groups 2 and 5, and 4 and 1 respectively); and the December and February data for three rotation groups. Only the January data will be represented by the full sample (rotation groups 1-4). In Wave 2, the persons originally interviewed in February are interviewed again in June for information on the months of February through May. In July, the March interviewees are asked to report data for March through June, and so on. After two interviews with the same rotation group, eight consecutive months of data are available. Then, if data collected in Wave 1 are used together with Wave 2 data, estimates for February through April can also be produced using all 4 rotation groups. Thus, to produce quarterly estimates for the full sample, it is necessary to work with more than one wave of data. In the same way, data corresponding to a calendar year can only be obtained by matching data from 4 consecutive waves of interviewing; e.g., data collected in interviews conducted from February 1985 (part of wave 1) through April 1986 (part of Wave 4) can be linked together to produce monthly data covering calendar year 1985.

In addition, this first section of the core collects much of the information on the receipt of income from various sources during the 4 month reference period. This includes income from government sources such as Aid to Families with Dependent Children, Supplemental Security Income, General Assistance, and Workmen's Compensation. Respondents are also asked about both Social Security and other retirement income including Railroad Retirement, pensions from company or union, and civil service retirments, as well as others. The receipt of miscellaneous sources of income such as alimony, child support, interest from savings, income for foster child care, and educational assistance is also identified. In addition, questions on major sources of noncash benefits such as food stamps, WIC (Women, Infants, and Children Nutrition Program), Medicaid, Medicare, and health insurance coverage are included in this section.

The second section of the SIPP core questionnaire collects information associated with wage and salary earnings. This section includes information on industry and occupation as well as hourly earnings for up to two jobs. Data are collected for two jobs held either concurrently or sequentially during the 4 month reference period. When more than two jobs occur (about 3 percent of the cases), data are collected for the two at which the person worked the most hours or the two most recently held jobs.

The third section of the core collects data on self-employment earnings and specific information about the kind of self-employment--whether it was incorporated, sole proprietorship, or partnership--and the profits and losses from the business. Again, space is provided for two self-employment jobs.

The fourth section is identified as the general amounts section. This section of the questionnaire collects monthly amounts received from the income sources identified in the first section. That is, the first section identifies the receipt of income during the 4 month reference period, while amounts of income received are collected in the fourth section of the questionnaire. Space is provided for amounts from up to six income sources. In Waves 1 and 2 of the 1984 Panel, no one had more than five separate income sources in this section of the questionnaire.

Survey Content

Each interview is planned to take about 30 minutes of a respondent's time and includes content that is divided into three main groups of questions. The substance of two of these groups should be essentially the same for each wave and for each panel. The third group of questions covers topics that will change in each wave of a panel. This will allow for the inclusion of some new content in each panel, although many of the topics will be repeated across all the panels. Each rotation group in a wave is administered the same set of questions although the reference period is different as explained above.

The first group of questions are control card items. The control card is a separate document from the questionnaire and serves several important functions. The control card is used to list every person residing at an address and to record basic social and demographic characteristics (age, race, sex, and so forth) for each person at the time of the initial interview. Some information relating to the housing unit or household also is collected; e.g., number of units in the structure, tenure, and so forth. The card is reused at subsequent interviews to record changes in characteristics such as age, educational attainment, and marital status, and to record the dates when persons enter or leave the household. Finally, during each interview, information on each source of income received and the name of each job or business is transcribed to the card so that this information can be used in the updating process at the next interview.

The second major group of questions form the core portion of the questionnaire, which is divided into 5 sections. The core set of questions is asked at the first interview and then updated in each subsequent interview. The first section of the core collects the basic labor force participation data for the 4 reference months. Respondents are asked to supply information on whether they had a job for all 4 months, and, if not, to answer a set of questions describing their activities when not at work. These categories include: "laid off," "looking for work," "not looking for work," "temporarily absent," as well as others. Labor force activity is collected on a weekly basis for all respondents with a job during the 4 month reference period. (See Ryscavage and Bregger (1985) for additional information on the labor force concepts used in SIPP.)

The variable topical modules are designed to satisfy the special programmatic needs of other Federal agencies. Time is set aside for variable modules in several waves to allow the flexibility to add content to meet special needs that develop as the survey continues. An example of a variable topical module is the child care topical module administered in the 1984 Panel. It was developed to obtain information about child care arrangements, such as who provides the care, the number of hours of care per week, where the care is provided, and the cost of the care. These data will be useful to other agencies because child care expenses are frequently deductible for program eligibility purposes. Variable topical modules may be repeated in subsequent waves or panels as necessary.

A wide variety of topics are covered under the aegis of the topical module concept. The breadth of these data ensure that SIPP will be a widely used and powerful data base serving multiple purposes. Appendix B contains a list of the fixed and variable modules scheduled for the 1984-1986 Panels and the address for obtaining a questionnaire with a specified module.

In addition to the data collected by the survey questionnaire, the content may be supplemented with administrative record data that are difficult for respondents to recall such as lifetime earnings and program benefit histories. To facilitate future linkages with administrative records, steps have been taken in SIPP to ensure that the social security number is obtained for as many persons as possible (see Appendix D).

Operational Procedures

Data collection operations are managed through the Census Bureau's 12 permanent regional offices. A staff of interviewers assigned to SIPP conduct interviews by personal visit each month. Self-response is required for each person 15 years old and older who is present at the time of interview and is obtained in about 65 percent of the cases. A proxy respondent is asked to provide information for those who are not available. Telephone interviewing occurs in about 5 percent of the cases to obtain missing information, to interview persons who will not or cannot participate otherwise, or to interview persons who have moved far outside the interviewing area. Most of the interviewing is completed during the first 2 weeks of a month.

The fifth and last section of the core questionnaire collects amounts of income earned from asset holdings. Asset sources include savings accounts, bonds, stocks, and rental property, as well as others. Information is collected for the 4 month reference period on both individual and joint recipiency.

See Appendix C for an explanation of income terms and a complete list of cash income types and examples of noncash benefits.

The third major question group consists of the various supplements or topical modules that are included in waves following the initial interview. The administration of a module is possible in Waves 2 through 8 (or 9 in 1984) because less time is required to update the core information after the first interview. Depending on the time available and length of the modules, more than one may be administered in the same wave. The topical modules cover areas that do not require examination every 4 months and may use a different reference period than the core questions. Some modules are assigned to only one wave of a panel, while other modules may be repeated in more than one wave. The modules provide a broader context for analysis by obtaining information on a variety of topics not covered in the core portion of the questionnaire. The module data may be analyzed independently or in conjunction with the control card items or core data. Frequently, a module is administered at the same time in concurrent panels so that the data may be combined to strengthen the analyses.

There are two types of topical modules: fixed and variable. The fixed topical modules are designed to be conducted on a regular basis to augment the core data. They are considered necessary to meet the survey's goals and objectives. Although the topics are "fixed," the questions in these modules may be modified from time-to-time to accomodate conceptual changes or to make improvements in collecting these data. An example of a fixed topical module is the annual "round-up" module on earnings and benefits. This module obtains wage and salary data from W-2 forms and estimates of annual self-employment for each appropriate person in the fifth and eighth interviews in each panel (sixth and ninth interviews in the 1984 Panel). Another fixed module administered at the same time obtains property income and tax-related information e.g., filing status and taxes paid, to allow the estimation of tax incidence, disposable income, and the simulation of tax policy alternatives.

A master list of identification numbers is used by the regional offices in monitoring the status of interviewing each month after Wave 1. The regional offices are responsible for ensuring that a completed questionnaire (or the reason for noninterview) is transmitted for each person assigned for interview. The production of reports and release of data are accelerated by processing questionnaires on a monthly basis.

Completed questionnaires undergo clerical edit before being entered into the Bureau's SIPP data processing system by keying performed in the regional offices. Upon entering this processing system, these data are subjected to a detailed computer check-in and preedit. Errors identified in this phase are corrected and computer processing continues. See Appendix F for a general description of clerical and computer editing procedures.

Two of the major steps of computer processing are the assignment of weights to each sample person and imputation for missing survey responses. The weighting procedures assure that SIPP estimates of the number of persons agree with independent estimates of the population within specified age, race, and sex categories. The procedures also assure close correspondence with monthly CPS estimates of households (see Appendix G for more information). Adjustments for unit nonresponses (a household in which no interview was completed) are made at the weighting stage. However, a nonresponse to a question in the survey is assigned a value in the consistency edit and imputation phase of processing. The imputation for missing responses is based on procedures generally referred to as the "hot deck" approach. This approach assigns values for nonresponses from sample persons who did provide responses and who have characteristics similar to those of the nonrespondents (see Appendix H for more information). The same approach is also used to fully impute data for individuals who are nonrespondents in otherwise cooperating households. The current weighting, editing, and imputation procedures are designed for cross-sectional data files only; however, research is underway on procedures for longitudinal data files.

Cross-sectional unit nonresponse rates for the first waves of the 1984 SIPP Panel show an improvement over the rates achieved in the Income Survey Development Program. Unit nonresponse (Type A) rates are the proportion of sample units

To carry out the longitudinal design of SIPP, interviewers collect information useful in locating persons (age 15 or older) who move; e.g., the name, address, and telephone number of someone who could supply the new address. A mover may be assigned to a different interviewer, possibly in a different regional office, who works closer to the new address. For cost reasons, personal visit interviews are only conducted at new addresses that are within 100 miles of a SIPP sampling area; telephone interviews are used otherwise. Persons who move into an institution, Armed Forces barracks, or outside the United States are not interviewed at the new location. When a sample person leaves an institution, interviewing resumes. (This procedure, however, was not implemented until the spring of 1985 .)

When an original sample person (those interviewed in the first wave) moves in with other people, all of the additional persons (age 15 or older) are interviewed in subsequent waves. Additional persons (age 15 or older) who move in with original sample persons are interviewed also. These additional persons are considered part of the sample and are interviewed only while residing with the original sample person(s). These provisions were adopted because most types of analysis using SIPP data will focus on the household and family situations of individuals. (See papers by Kalton and Lepkowski (1985) and Jean and McArthur (1984) for further discussion of following movers.)

Each person is assigned a unique identification number at the time he/she enters the sample. Once assigned, the person's ID rarely changes. The identification numbering system was created to allow the linkage of individual data from one point in time to another and the linkage of each person's data to the household or family to which he/she belonged during any month in the panel. Control card information, such as "date entered the household" and "date left the household," is used to group persons into households for each survey month during processing. Therefore, it is possible to link data from all persons ever associated in the same household at any time throughout the 2 1/2-year duration of a panel. This facilitates the construction of household income estimates based on the actual composition of households during the measurement period. Further explanation and a brief example of the use of the numbering system are provided in Appendix E. (See also Jean and McArthur, 1984.)

releasing the main statistics. Eventually, reports will be issued that examine detailed data from the topical modules in relation to income and program participation. For example, a report might focus on the relationship between disability and earnings or between health insurance coverage and labor force participation. These reports may also focus solely on the material covered in a topical module such as work history or migration.

When the SIPP longitudinal data base linking the waves together is completed, reports examining trends and changes over time will be prepared. For example, they may include analyses of the dynamic aspects of the labor force or the effect of changes in household composition on economic status and program participation.

Six kinds of longitudinal reports have been proposed for consideration: 1) economic profile reports, presenting yearly aggregates of monthly data on individuals; 2) comparative profile reports, presenting comparisons of yearly aggregates of monthly data on individuals; 3) transition reports, providing changes in income and program participation status between two points in time; 4) multiple transition reports, providing patterns of labor force, income and program participation activity, and the number of times or spells in a given status as well as the duration of each spell; 5) longitudinal family and unrelated individual reports, presenting the characteristics of longitudinal family units (as defined in SIPP--see McMillen and Herriot, 1984 for more information), as well as the characteristics of unrelated individuals who complete the sample universe; 6) special event reports, providing data related to a unique event such as marriage, the birth of a child, a return to school, or a move. These events are "special" in the sense that the economic situation of the person, both prior to and after the event, is of interest.

SIPP Public-Use Microdata Files

A variety of public-use microdata data files also will be released. These files consist of unaggregated records for individual survey respondents; however, the individuals cannot be identified in accordance with the Census Bureau's confidentiality policy. Several files containing core data collected in a full wave are available now. Appendix J contains the schedule for releasing core files and the address for obtaining files.

occupied by persons eligible for interview for which no interview was completed. Refusal households account for the largest part of this rate which also includes no one at home and temporarily absent households. After 5 interviews, the Type A rate for SIPP was 13.4 percent; for ISDP it was 14.3 percent. The 1985 Panel's first wave noninterview rate is somewhat higher than the rate for the first wave of the 1984 Panel, but it is still lower than the corresponding ISDP rate. Information concerning nonresponse rates for selected questions or "items" on the questionnaire also is available in several papers (Coder and Feldman, 1984; Lamas and McNeil, 1984; McMillen and Kasprzyk, 1985.) and in the appendix on data quality appearing in the SIPP quarterly report series (P-70).

SIPP Publications

In September 1984, the Census Bureau released the first SIPP data report titled Economic Characteristics of Households in the United States: Third Quarter 1983. Examples of data included in this report are shown in Appendix I. This report is part of a series of quarterly reports (Series P-70) that show average monthly labor force activity, income, and program participation statistics for one-quarter of the year. The procedures used to compute monthly averages treat each month of the quarter as an independent observation. The monthly average estimate is computed by calculating the number of interest (e.g., the number of persons receiving food stamps) for each individual month, summing the three estimates, and dividing the sum by three. The estimates are based on data from the full sample (all 4 rotation groups) for all 3 months of the quarter, except at the beginning of a panel when the sample is being phased-in.

A quarterly report is released approximately 1 year after the end of the reference quarter. This delay results in part from the monthly interviewing scheme which requires sample data from 6 months of interviewing to produce a quarterly estimate (see Figure 2). The schedule for releasing quarterly reports and the address to obtain copies of these reports are contained in Appendix J, and a list of table titles appearing in this series of reports is in Appendix K. Tables which show monthly averages for a calendar year period are planned also.

Several other types of reports being considered for publication will take longer to produce because the data are not ready as quickly. For example, the topical module data are processed after the core information to prevent delays in

metropolitan areas to be specifically identified on the files. To also reduce the risk, income from each source is topcoded at $100,000 per year. A few other data items are regrouped (i.e., recoded) or suppressed to help maintain the confidentiality of respondents. Technical documentation including a data dictionary will be prepared for each public-use file and may be purchased (see address in Appendix J) to determine exactly what data are available.

The wave files are available for purchase by the public in both a relational and rectangular format. The relational structure is fairly complex and difficult for nonprogrammers to use. There are eight types of records: sample unit, household, family, person, wage and salary, self-employment, asset income, and other income sources. The relationships between the records are expressed by pointers on each record. For example, a person record would have a pointer noting the household to which he/she belongs. Correspondingly, the household record would have pointers to the records of each person who is a household member. Therefore, the format of each record is variable.

Some users, particularly those who want to use statistical software packages, may prefer the rectangular structure. A rectangular file has one type of record of consistent length and fixed format throughout. Each logical record for a sample person includes information on the household and family of the person during each month of the reference period, as well as characteristics of the person and each source of income received.

The relational and rectangular wave files are fully edited, imputed, and weighted for use in cross-sectional analyses. Users can create a longitudinal file by matching two or more of these wave files; however, the user will have to develop longitudinal edits, allocation values, and weights for these merged files. The Census Bureau does intend to make edited, imputed, and weighted longitudinal files available to the public at a later date. Files containing core and topical module data for a specified wave also will be prepared for public release. The first two topical module files, from Waves 3 and 4 of the 1984 Panel, should be available in February and January 1986, respectively.

Virtually all of the collected data is contained on the public-use files. However, certain questionnaire items are collapsed or deleted from the rectangular file to provide a more workable file format for data users. In addition, other restrictions are applied to both types of files to maintain the confidentiality of individual respondents. For example, name, address, social security number and other unique identifiers are eliminated from the files. The risk of disclosure is also reduced by only identifying geographic areas that contain at least 250,000 residents within the sampled area. This limitation still allows residence in most States and many

decreased due to attrition. In FY 1985, budget constraints required a reduction in the sample size, which was about 19,000 interviewed households then. Beginning in March (1985), about 850 interviewed households were deleted from each of the 4 rotation groups. After the reduction, about 15,600 households remained in the sample. Unless there are further reductions, the 1984 Panel will continue at approximately this level until it ends in August 1986.

The 1985 and Subsequent Panels

The sample for the 1985 Panel and subsequent panels is in 230 Primary Sampling Units (PSUs) of which 86 are self-representing and 144 are nonself-representing. The PSUs were selected so that each state has a sample area with certainty. The sample design involved multiple stages of selection.

The main sampling frame for selecting living quarters was the 1980 decennial census address list, which was updated for new construction. The overall probability of selection for all living quarters in the SIPP sample was not the same, but the probabilities were very close. To select an efficient sample, a systematic sample of segments of living quarters was selected within each sample PSU. Some segments had an expected size of 2 living quarters per segment; other segments had an expected size of 4 living quarters.

Before the 1985 Panel sample was implemented in February 1985, budget constraints required a reduction in the planned sample size. Therefore, the 1985 Panel started with a sample of about 17,800 designated living quarters resulting in an interviewed sample size of approximately 13,300 households. Subsequent panels are expected to start at approximately this size also.

APPENDIX A

SIPP Panel Sample Selection Procedures

The SIPP universe is the noninstitutionalized resident population living in the United States. This population includes persons living in group quarters, such as dormitories, rooming houses, and religious-group dwellings. Crew members of merchant vessels, Armed Forces personnel living in military barracks, and institutionalized persons, such as correctional facility inmates and nursing home residents, are not eligible to be in the survey. Similarly, United States citizens residing abroad are not eligible to be in the survey. Foreign visitors who work or attend school in this country and their families are eligible; all others are not eligible to be in the survey. With these qualifications, persons who are at least 15 years of age at the time of the first interview are eligible to be in a panel of the survey.

The 1984 Panel

The 1984 SIPP sample is located in 174 areas, called Primary Sampling Units (PSUs), comprising 450 counties (including one partial county) and independent cities. Usually a group of counties rather than only a single county is used as a PSU. The PSUs were combined into strata, and one PSU was selected from each stratum. The PSUs with the largest populations were treated as separate strata and included in the sample with certainty; they are called self-representing (SR). The remaining 129 PSUs are called nonself-representing (NSR) and were selected with probability proportional to size. The sample consists mostly of clusters of two to four living quarters, called segments. The bulk of the segments were systematically selected from lists of addresses prepared for the 1970 decennial census. The sample was updated to reflect new construction since 1970.

At the start of the 1984 Panel, 26,000 living quarters were designated for interview. Approximately 5,000 households subsequently were eliminated from the sample because the units were demolished, unoccupied, or ineligible for some other reason. Therefore, the eligible sample size for interviewing was about 21,000 households. The eligible sample size increased slightly with each subsequent wave because of the procedures for interviewing new household members who joined movers. Overall, however, the number of interviewed households gradually

help respondents supply the information. The short wave is wave 2 in the 1984 and 1985 Panels and wave 3 in the 1986 Panel. For most persons using wave files for cross-sectional purposes this variation in interviewing schedule will have little effect, assuming it has no effect on rotation group bias. However, some users of wave files will have to concern themselves with this problem. In particular, persons building longitudinal files from two or more wave files and those persons constructing quarterly or other time-based files must be wary. A two-wave longitudinal file that includes the short wave will have only 75 percent of the sample. Similarly, any longitudinal file containing the short wave will have 25 percent of the sample with one fewer interview.

In addition to these operational differences, there are differences in the content of each panel. Figure B-1 outlines the content of the 1984, 1985, and 1986 Panels. A summary of the content of the 1984 Panel is also available in Frankel (1985).

Some of the topical modules called variable topical modules, are designed to have content which varies from panel-to-panel. It is the intent that these modules be used to serve the needs of the federal policy community. In the 1984 and 1985 Panels variable sections were distributed across all waves containing topical module questions. For the 1986 Panel these questions were consolidated into wave 3 and wave 6 modules. In addition, several of the previously separate topical modules have been combined into one restructured module in the 1986 Panel; namely, the modules called Welfare History, Work History, Disability, Education and Training, Migration History, Fertility History, Marital History, and Household Relationships are all parts of the Personal History topical module. The Personal History topical module is considered to be fixed; i.e., it will be repeated in subsequent panels. The specific content of a module listed in Figure B-1 may be obtained by writing for a copy of the appropriate wave questionnaire: Income Surveys Branch, DSD, Room 3339, FOB 3, Bureau of the Census, Washington, DC 20233, or call (301) 763-2063.

APPENDIX B

Differences Between SIPP Panels

Minor modifications in both the design and content of SIPP have been made between the 1984, 1985, and 1986 Panels. It is likely that changes will continue to be made to future panels as the Census Bureau develops more efficient methods and improved questions. In addition, some variation in content is part of the basic design of the survey. This appendix outlines the changes in design first and then the changes in content.

The 1984 Panel of the SIPP is slightly different from subsequent panels in the size, numbering, and scheduling of interviews. First, the 1984 Panel began with approximately 26,000 designated households while all other panels are designed to begin with approximately 18,000 households (see Appendix A for actual panel sizes). The reduction in sample size occurs because of sample redesign research which indicated that acceptable coefficients of variation for impotant survey variables can be achieved at the reduced sample size.

Second, the 1984 Panel began in October 1983 to get the survey started as soon as possible. This early start extended the 2-2/3 years normally covered by a panel by several months. As a result the 1984 Panel has 9 waves while all others have only 8. Both the 1985 and 1986 Panels begin in February, which is the proposed starting month for all future panels.

Third, each panel has one rotation group which is interviewed one less time than the other three during the life of the panel. In the 1984 Panel, one rotation group receives nine interviews, and three receive eight interviews. In the 1985 and 1986 Panels, one rotation group receives eight interviews and three rotation groups receive seven interviews. This occurs because one wave in each panel covers only three rotation groups. The shortened wave was created to align the initiation of waves across panels--a procedure that simplifies field operations. In addition the short wave is necessary to get the interviewing on a cycle that allows the questions concerning taxes and annual income to be asked in May through August each year when W2 and Internal Revenue Service (IRS) forms are available to

Figure 8-1. SURVEY OF INCOME AND PROGRAM PARTICIPATION
TOPICAL MODULE SCHEDULE
1984/1985 Panels

INTERVIEW DATES	1984 PANEL Wave	1984 PANEL Fixed Topical Module	1984 PANEL Variable Topical Module	1985 PANEL Wave	1985 PANEL Fixed Topical Module	1985 PANEL Variable Topical Module
Oct 83-Jan. 84	1	None	None			
Feb 84-Apr 84	2	None	None			
May 84-Aug 84.	3	Health and Disability Work History Education History				
Sept 84-Dec. 84	4	Assets Liabilities	Pension Plan Coverage Characteristics of Job from which Retired Retirement Plans and Expectations Housing Costs and Conditions Energy Usage			
Jan 85-Apr 85	5		Child Care Arrangements and Expenses Welfare History Child Support Support for Non-Household Members Reasons for Not Working Reservation Wage Work Related Expenses	1	None (Feb. 1985-May 1985)	None
May 85-Aug 85	6	Annual Income Taxes Employee Benefits Educational Financing and Enrollment	Training Questions (ETA)	2	None (June 1985 - Aug. 1985)	
Sept. 85 Dec. 85	7	Assets Liabilities	Pension Plan Coverage Update	3	Assets Liabilities	
Jan. 86-Apr 86	8	Marital History Fertility History Migration History	Household Relationships Support for Non-Household Members Work Related Expenses	4	Marital History Fertility History Migration History	Household Relationships Support for Non-Household Members Work Related Expenses
May 86-Aug 86	9	Annual Income Taxes Individual Retirement Accounts Educational Financing And Enrollment		5	Annual Income Taxes Individual Retirement Accounts Educational Financing and Enrollment	
Sept 86-Dec 86				6		Child Care Arrangements Child Support Agreements Support for Nonhousehold Members Job offers Health Status and Utilization of Health Care Services Long-Term Care Disability Status of Children
Jan 87-Apr 87				7	Assets Liabilities	Pension Plan Coverage Lump Sum Distributions from Pensions Plans Characteristics of Job from which Retire Characteristics of Home Financing Arrangements
May 87-Aug 87				8	Annual Income Taxes Individual Retirement	

A few changes have been made to the core questions since the 1984 panel was introduced. In the 1985 Panel, core questions on education were modified to capture enrollment status on a wave-to-wave basis. In addition, the exit codes for persons leaving sample households were expanded to provide more detailed reasons. In the 1986 Panel, entry codes were expanded to provide symmetry with the exit codes. In addition, a set of questions aimed at identifying discouraged workers were dropped. The field and analysis staffs identified those questions as excessively burdensome. The discouraged worker questions were replaced by questions aimed at providing information to allow more strict comparisons of CPS and SIPP labor force data.

APPENDIX C

Explanation of Income Concepts and Noncash Benefits

Monthly income. For each person in the sample who is 15 years old and over, questions are asked about the amount of money (cash) income received each month from a variety of separately identified sources. The income amounts represent the amounts actually received during the month, before deductions for income and payroll taxes, union dues, Part B Medicare premiums, etc. While the income amounts from most sources are recorded monthly for the 4-month reference period, property income amounts such as interest, dividends, rental income, etc., are recorded as totals for the 4-month period. These totals are distributed equally between months of the reference period for purposes of calculating monthly averages.

Cash income. Cash income is the sum of all income received from any of the sources listed in Figure C-1. Rebates, refunds, loans, capital gain or loss amounts from the sale of assets, and interhousehold transfers of cash such as allowances are not included. Accrued interest on Individual Retirement Accounts, KEOGH retirement plans, and U.S. Savings Bonds is excluded also. Lump-sum or one-time payments, such as inheritances or insurance settlements, are included, which is the main difference from the March CPS cash income definition. Cash income is divided into 3 primary sources: 1) earnings from employment, 2) income from assets (property income), and 3) other cash income.

Earnings income. The SIPP income definition includes three types of earnings: wages and salary, nonfarm self-employment, and farm self-employment. Earnings from all such sources are included for each person. The definition of nonfarm self-employment and farm self-employment is not based on the net difference between gross receipts or sales and operating expenses, depreciation, etc. The monthly amounts for these income types are based on the salary or other income received from the business by the owner of the business or farm during the 4-month period.

Income from assets (Property income). See Figure C-1 for a list of the types of income included in this concept.

Other income. See Figure C-1 for a list of the types of income included in this

Figure B-1 SURVEY OF INCOME AND PROGRAM PARTICIPATION
TOPICAL MODULE SCHEDULE - Continued
1986 Panel

INTERVIEW DATES	WAVE	1986 PANEL Fixed Topical Module	Variable Topical Module
Feb. 86-May 86	1	None	None
June 86-Sept. 86 *	2	Fertility History Marital History Migration History Recipiency History Employment History Work Disability History Education and Training History Family Background Household Relationships	
Oct 86-Dec 86	3		Child Care Arrangements Child Support Agreements Support for Nonhousehold Members Job Offers Health Status and Utilization of Health Care Services Long-Term Care Disability Status of Children
Jan 87 - Apr 87	4	Assets Liabilities	Pension Plan Coverage Lump Sum Distributions from Pension Plans Characteristics of Job from which Retired. Characteristics of Home Financing Arrangements
May 87-Aug 87	5	Annual Income Taxes Individual Retirement Accounts Educational Financing and Enrollment	
Sept. 87-Dec. 87	6		Child Care Arrangements Child Support Agreements Support for Nonhousehold Members Work Related Expenses Housing Costs Energy Usage
Jan. 88-Apr. 88	7	Assets Liabilities	
May 88 - Aug 88	8	Annual Income Taxes Individual Retirement Accounts Educational Financing and Enrollment	

* These modules are collectively identified as the Personal History Topical Module.

APPENDIX D

Links to Administrative Records and the Validation of Social Security Numbers

The SIPP data system has always been thought of as a combination of data from administrative records and household surveys. This reduces respondent burden by using other data sources for difficult-to-obtain information. Interview responses can be supplemented by information from program files such as the earnings and benefit records of the Social Security Administration. This allows, for example, analysis of the long-term impact of various Social Security benefit formulas.

To make these linkages accurate, social security numbers (SSN) are required for sample individuals. The SSN is obtained for each household member in SIPP and recorded on the control card. It is identified as a critical survey data item requiring completion to make the interviewers aware of its importance. These efforts should result in improved accuracy of the survey-reported social security numbers. These numbers are then verified and corrected to maximize the number of accurate linkages to other record systems.

The verification and correction process builds on the work of the development program in which SSNs were obtained and verified for more than 95 percent of the sample (see "Social Security Number Reporting, the Use of Administrative Records, and the Multiple Frame Design in the Income Survey Development Program" by D. Kasprzyk in David, M. (ed.), 1983). At the conclusion of each month's interviewing during the first wave of a SIPP panel, a special extract file is prepared by the Census Bureau for the Social Security Administration. This file contains a small number of key variables (SSN, name, date of birth, sex) for all original sample persons who report a SSN, including children, in a format appropriate for machine validation. Persons who report that they have a number but cannot supply it or that they do not have a number are handled separately in a clerical (manual) procedure. Persons who refuse to provide a SSN are not included in the search process. The Social Security Administration identifies (by machine validation) incorrectly reported numbers then also clerically resolves these cases along with cases with a missing SSN. This work is completed by the fourth wave interview, at which time a field followup is conducted to obtain missing SSNs (provided they are not "refusals") and to reconcile inconsistencies in SSN or demographic data generated by the computer match or the clerical resolution.

Figure C-1. Income Sources Included in Monthly Cash Income

Earnings from Employment

Wages and salary
Nonfarm self-employment income
Farm self-employment income

Income from Assets (Property Income)

Regular/passbook savings accounts in a bank, savings and loan, or credit union
Money market deposit accounts
Certificates of Deposit or other savings certificates
NOW, Super NOW, or other interest-earning checking accounts
Money market funds
U.S. Government securities
Municipal or corporate bonds
Other interest-earning assets
Stocks or mutual fund shares
Rental property
Mortgages
Royalties
Other financial investments

Other Income Sources

Social Security
U.S. Government Railroad Retirement
Federal Supplemental Security Income (SSI)
State Administered Supplemental Security Income
State unemployment compensation
Supplemental Unemployment Benefits
Other unemployment compensation (Trade Adjustment Act benefits, strike pay, other)
Veterans' compensation or pensions

Other Income Sources-Continued

Black lung payments
Worker's compensation
State temporary sickness or disability benefits
Employer or union temporary sickness or disability benefits
Payments from a sickness, accident or disability insurance policy purchased on your own
Aid to Families with Dependent Children (AFDC, ADC)
General assistance or General relief
Indian, Cuban, or Refugee Assistance
Foster child care payments
Other welfare
Child support payments
Alimony payments
Pensions from a company or union
Federal Civil Service or other Federal civilian employee pensions
U.S. Military retirement pay
National Guard or Reserve Forces retirement pay
State government pensions
Local government pensions
Income from paid-up life insurance policies or annuities
Estates and trusts
Other payments for retirement, disability, or survivors
G.I. Bill/VEAP education benefits
Income assistance from a charitable group
Money from relatives or friends
Lump sum payments
Income from roomers or boarders
National Guard or Reserve pay
Incidental or casual earnings
Other cash income not included elsewhere

Noncash benefits. The major sources of noncash benefits are: Food Stamps, Special Supplemental Food Program for Women, Infants, and Children (WIC), Low-Income Home Energy Assistance, Medicaid, Medicare, Free or reduced-price school lunches and breakfasts, Public or subsidized rental housing, and health insurance coverage.

APPENDIX E

The SIPP Identification Numbering System *

The SIPP identification scheme is designed to uniquely identify individuals in each wave. It provides a means of linking data for the same individuals over time and it is useful for grouping individuals into unique households each wave. The identification scheme does not number each "different" household viewed in a longitudinal way over time. However, all households newly formed after Wave 1 due to splits can be linked to the original Wave 1 household through the identification numbering system.

Demographic information identifying family relationships, handling changes in relationships, etc., are not discussed in this appendix.

The various components of the operational identification scheme are:

PSU number	-3 digits
Segment number	-4 digits
Serial number	-2 digits
Current Address ID	-2 digits
Entry address ID	-2 digits
Person number	-3 digits

Census Bureau-Washington staff assign PSU, segment, and serial numbers during sample selection. The three-digit PSU number identifies a county or group of counties and is the same number used by other census surveys such as the Current Population Survey (CPS) and the National Crime Survey (NCS). As a sample of segments is drawn from a PSU, the segments are uniquely numbered within each PSU, using a four-digit number. A two digit serial number is assigned to each unit in a segment.

* Anne Jean, Demographic Surveys Division, is the author of this appendix.

Social security numbers of persons who enter the sample after Wave 1 (because they start living with original sample people) are validated at the start of the next panel. For example, information on new panel members (nonsample persons) from Waves 2 through 5 of the 1984 Panel was held and submitted for computer validation with Wave 1 of the 1985 Panel. Likewise, information on nonsample persons from Waves 6 through 8 of the 1984 Panel and Waves 2 through 4 of the 1985 Panel will be held and submitted for computer validation with Wave 1 of the 1986 Panel.

At this time, definite plans have not been made for accessing Federal files either for appending program information to the survey household records or for evaluating the quality of the SIPP data by comparing the survey-reported data with administrative records data. A Census Bureau committee is, however, studying the uses of administrative records in the SIPP and will make recommendations for future work. One application, not related to survey data evaluation, that has been discussed extensively within the Census Bureau is a possible match of SIPP survey data with data about the employers of the sample respondents. See Haber (1985) for a discussion of some potential uses of a SIPP file that is linked to micro-level establishment and enterprise data from the Bureau's economic censuses and other sources.

the roster in Wave 2; and so forth. The first digit indicates the wave the person enters the survey. The five-digit number consisting of entry address ID and person number generally is not changed or updated in subsequent interviews, even when a person moves. In rare instances, where persons from different SIPP households (i.e., households having different PSU, Segment, Serial Numbers) move in together, these numbers are changed. This is handled on a case-by-case basis.

Thus, the fourteen-digit combination of PSU, segment, serial, entry address ID, and person number uniquely identifies each person in the SIPP survey, and can be used to link data for the same persons across waves. The PSU, segment, serial, and current address ID code uniquely identifies each household in any given wave, and the PSU, segment, and serial number can link all households in subsequent waves back to the original Wave 1 household.

To illustrate the SIPP identification numbering system, consider the following example (see Figure E-1). Assume that a three-person household exists in Detroit in Wave 1 and it consists of a husband, wife, and an 18 year-old son. In Wave 2, the son moves and sets up a household with a friend in a nearby city. The husband and wife are still living at the Wave 1 address. In Wave 3, the wife retires and moves to Florida to live with her sister, while the husband moves to California to live with a family friend and her two children. In this example, note that the PSU-segment-serial portion of the identification number is constant and that the entry and person numbers do not change once they are assigned. However, the current address ID does change when a member of the original household moves to a new address.

Office staff in the twelve regional offices also assign a two-digit serial number to any additional units discovered during the field listing operation. In Wave 1, the combined PSU, segment, and serial number uniquely identifies one sample address and all of the persons living at that address are assigned the same set of 9 digits. A person's PSU, segment, and serial number is not changed in later waves, even if the person moves. This combination of 9 digits is the link to the original sample unit. All additional persons who later live with an original sample person receive the same 9 digits as the original sample person. A scrambled version of these 9 digits appears on the public-use file and is called the Sample Unit ID.

The regional office staff also assigns a two-digit current address ID code. It is added to provide a means for identifying more than one unique household represented by the same PSU, segment, and serial number. The first digit of the current address ID code indicates the wave that an address is first assigned for interview. The second digit is used to sequentially number addresses for households that split into two or more households as a result of a move to different locations by the original sample persons. The Wave 1 current address ID code for all sample addresses is 11, since no splits occur until after Wave 1. When a SIPP sample person moves to new address, the office staff assigns a new current address ID code to the new address. Previous address ID codes assigned to the old address are deleted from the processing system if no SIPP sample persons remain at the previous address. Thus, the combination of PSU, segment, serial number and current address ID code uniquely identifies each sample address for any given wave. As only one sample household is associated with a sample address, this combination provides unique household identifiers for a given wave.

The person ID is a five-digit number consisting of two parts: entry address ID and person number. It is assigned by the interviewer to each sample person during the initial listing of the household roster. Additional persons who join SIPP households after the initial interview are assigned their person IDs at the time they are added to the roster. The interviewer uses the current two-digit address ID code for that address as the person's entry address ID. It indicates the address of the household at the time the person entered the sample. Next, the interviewer assigns a three-digit person number to each person. Numbers 101, 102, etc., are assigned in Wave 1; 201, 202, etc., are assigned to persons added to

APPENDIX F

SIPP Edits

Clerical Edit. Performed by regional office staff on a sample of questionnaires from each interviewer to detect omissions, errors, or misunderstandings.

ENTREX Edit. Performed in the regional offices when the control cards and questionnaires are keyed to ensure that: 1) the data are keyed in the proper sequence; 2) certain data are present (.e.g., control number, name, relationship to householder description); and 3) a limited set of numeric items, mostly on the control card, are within a specified range. Data failing edit are rekeyed.

Preedit. Performed on all core data transmissions as they are received in a central location from the regional offices. (Topical module data are not preedited.) The preedit is similar to ENTREX editing, but it is more extensive. The core data preedit program:

1. Ensures receipt of all expected cases.

2. Range checks the data.

3. Checks for completeness on specified items.

4. Performs limited consistency edits.

5. Checks for possible keying errors in codes used to identify data items on a computer record.

Errors identified in this edit are described on a reject listing for the regional offices. The office staffs resolve the problems by looking at the documents or contacting the interviewers; the corrections are then keyed.

Figure E-1. Illustration of the SIPP Identification Numbering System

	PSU	SEGMENT	SERIAL	ADDRESS ID	ENTRY ID	PERSON
WAVE 1						
Husband	101	1111	01	11	11	101
Wife	101	1111	01	11	11	102
Son	101	1111	01	11	11	103
WAVE 2						
			Husband-Wife Household			
Husband	101	1111	01	11	11	101
Wife	101	1111	01	11	11	102
			Son-Friend Household			
Son	101	1111	01	21	11	103
Friend of son	101	1111	01	21	21	201
Wave 3						
			Husband's Household			
Husband	101	1111	01	31	11	101
Friend of husband	101	1111	01	31	31	301
Child 1	101	1111	01	31	31	302
Child 2	101	1111	01	31	31	303
			Wife's Household			
Wife	101	1111	01	32	11	102
Sister	101	1111	01	32	32	301
			Son's Household			
Son	101	1111	01	21	11	103
Friend of son	101	1111	01	21	21	201

APPENDIX G

Overview of Cross-sectional Weighting for SIPP*

The final monthly weight for each SIPP sample person is the product of the four factors described below.

The first factor is a weight appropriate for unbiased estimation of levels and proportions in the absence of nonresponse. For the first wave, this factor is simply the inverse probability of selection. With few exceptions, this factor will be uniform across cases. For subsequent waves, this factor decreases for housing units that have members who were not part of the original sample (excluding newborn babies and persons returning from institutions, overseas, or the Armed Forces).

The second factor is a correction for household nonresponse. For the first wave of interviewing, information on census region, residence status, race of reference person, tenure (own or rent), and household size is used to make this correction. (Interviewers make a concerted effort to obtain this information for all households whether or not they respond). For the second and subsequent interviews, the household noninterview adjustment factors must account for household refusals and temporarily absent households. They also must account for households whose members all move to an unknown address or to a place more than 100 miles from the nearest SIPP sampling area, called a Primary Sampling Unit (PSU) and cannot be contacted by telephone. Information from the previous wave -- race of reference person, household type, education level of the reference person, household income sources, household assets, tenure (own or rent), public housing or rent subsidy, and household size -- is used to calculate the household noninterview adjustment factor.

The third factor is intended to reduce the between PSU variance. This factor is applied to sample cases in nonself-representing (NSR) areas. It is achieved by post-stratifying within region by metropolitan status and race in 1980.

Person and Family Edit. Creates family records using codes assigned to persons on the basis of their relationship to the householder (i.e., the person who owns or rents the house). Key variables used in weighting, such as age, race, and sex, are imputed at this point.

Consistency Edit. Ensures the logical consistency of responses recorded for persons, families, and households. Consistency is examined 1) within and between sections of the questionnaire, and 2) between the questionnaire, control card, and other documents.

Control Card Edit. Edits data normally collected on the control card and imputes data if certain characteristics of the sample unit are missing. Among these characteristics are the number of units in the structure, tenure, and information about public and subsidized housing.

I&O Edit. Edits and imputes when necessary, three-digit industry and occupation codes which are clerically derived from information entered in the employer and self-employment sections of the questionnaire.

APPENDIX H

Cross-Sectional Imputation System for Item Nonresponse

The imputation for item nonresponse is made using the Census Bureau's traditional sequential hot-deck procedure. The data file is sorted to provide some control for geographical location and processed sequentially. The cells of the hot-deck matrices are then loaded with data from respondents (one value per cell). The value existing in a specified cell is assigned to a nonrespondent with characteristics identical to those defining the hot-deck cell.

The hot-deck matrices must be preloaded with values prior to any imputation. This preloading is a two-step operation. First, initial or "cold-deck" values for each cell are supplied. Second, "live" values from donors are loaded into the matrix by allowing an initial run of the imputation programs that store data in the hot deck but perform no actual imputations. This procedure has been used in both the March CPS and 1979 Research Panel of the Income Survey Development Program. The imputation system also produces detailed diary information on the performance of the hot decks and places imputation flags on the microdata file for each item imputed. A flowchart of the imputation system is displayed on the next page (Figure H-1).

The fourth factor accomplishes several goals simultaneously. It reduces the mean square error of estimates of characteristics of persons correlated with age, race, or sex. Independent current demographic estimates of population by age, race, and sex are used for this purpose. In addition, monthly estimates of type of household (marital and family status of householders by race and sex) from the Current Population Survey (CPS) are used to reduce variances on estimates related to these controls. Under this procedure, the monthly estimates are consistent with the March CPS publications. To satisfy these mildly conflicting objectives while keeping the number of husbands equal to the number of wives and the number of householders equal to the number of households, an iterative raking procedure is used. For the 1985 Panel, Hispanic controls by age and sex also will be used in producing the fourth factor to reduce the mean square error of the Hispanic estimates. Use of such controls for the 1984 Panel is still under consideration.

*Raj Singh, Statistical Methods Division, is the author of this appendix.

Data Examples Taken from Economic Characteristics of Households in the United States: Third Quarter 1983

Monthly Household Income

A household's cash income is the single most commonly used measure of economic well-being. It is often compared to other household characteristics such as race, regional location and age of the householder.

Chart A shows that the median monthly cash income for the nation's 83.1 million nonfarm households ("All" on Chart A), was $1,670 for the third quarter of 1983. White households during this period had a median monthly income of $1,750 while the median monthly income for Black households was $1,080 and $1,230 for Spanish-origin households.

When median monthly cash income is compared by regional location of the household, Chart B, the households in the West and Northeast are the highest with incomes almost the same at $1,800 for the West and $1,770 for the Northeast. The Midwest is a little lower at $1,650, with the South the lowest at $1,550.

Age is another household characteristic which is compared by median monthly income. The age used here is that of the householder. Chart C shows median monthly income rising as the age of the householder increases, peaking in the category of 45 to 54 years, with median monthly income of $2,340. The median monthly income then declines to $1,810 in the 55 to 64 age group and to $950 in the 65 years old and over group. This decline results from a variety of reasons including retirement, increased work disabilities and smaller sized households with fewer people contributing to a household's income.

Median Monthly Cash Income for Nonfarm Households

(Third Quarter 1983)

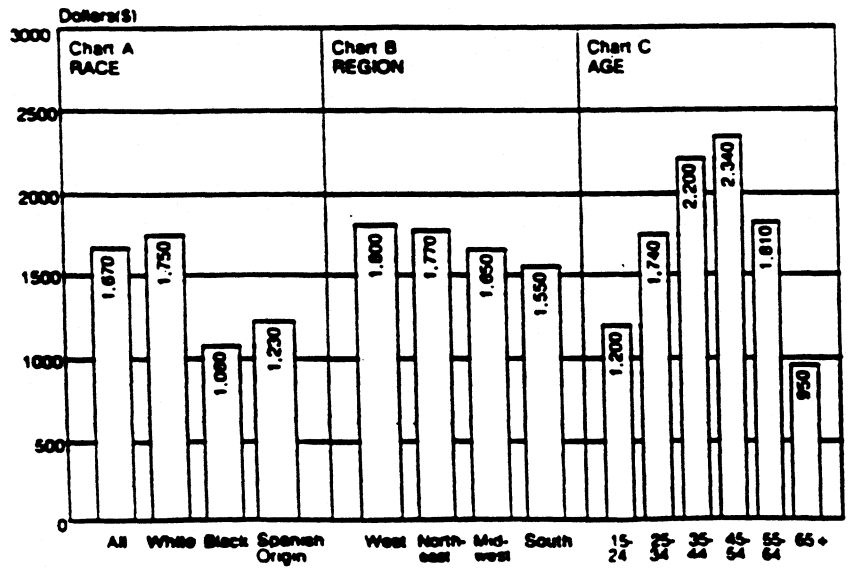
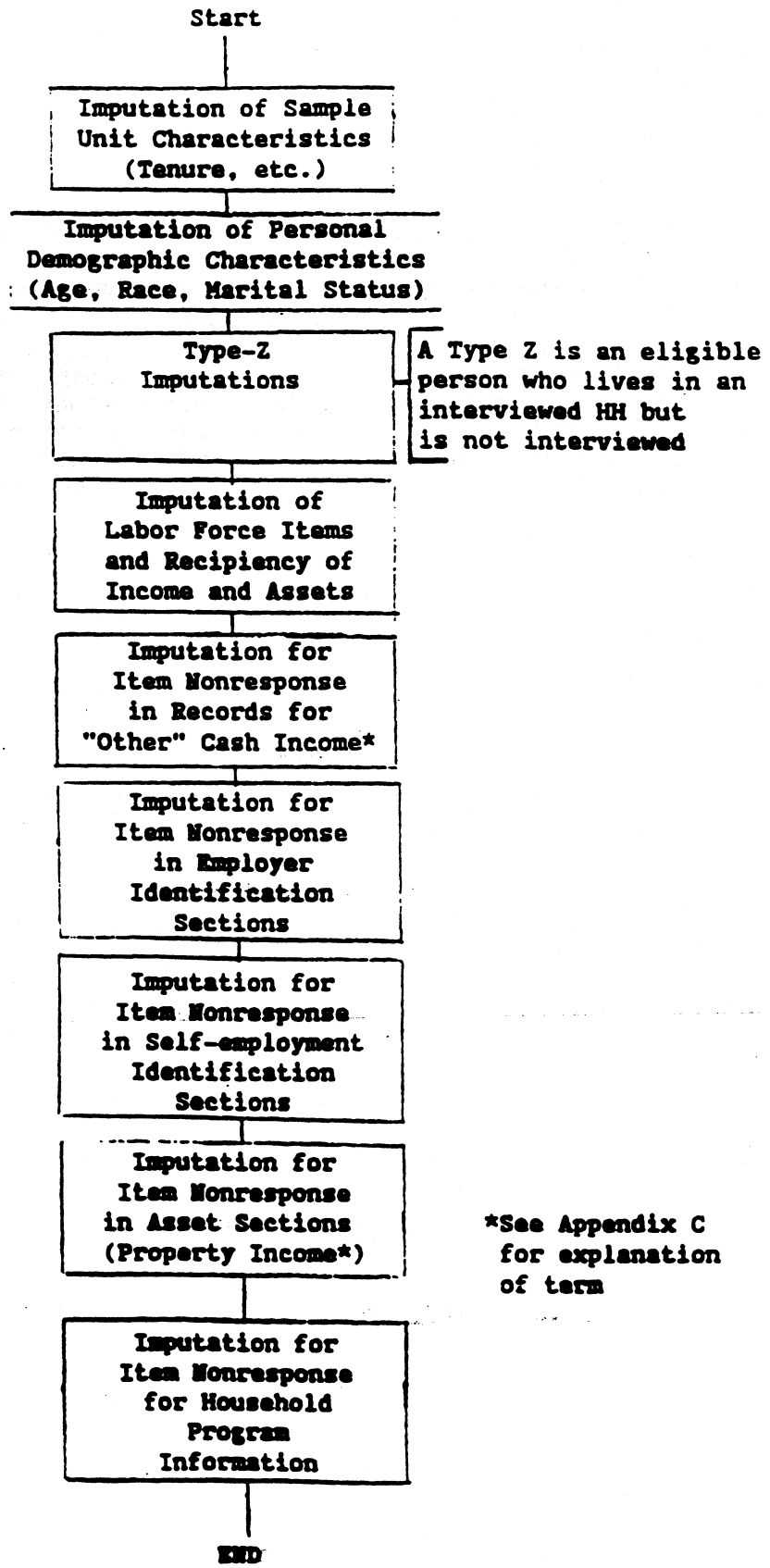
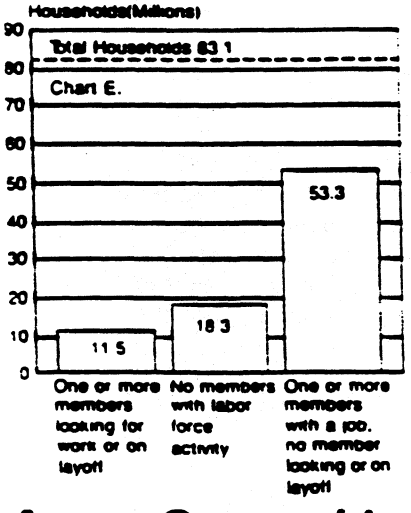


Figure H-1. Cross-Sectional Imputation System Flowchart



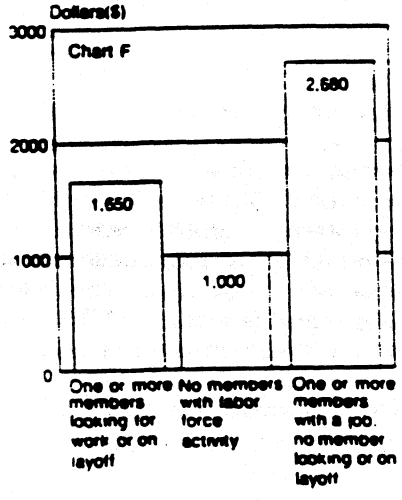
Households, by Labor Force Activity

(Third Quarter 1983)



Mean Monthly Cash Income of Nonfarm Households by Labor Force Activity

(Third Quarter 1983)

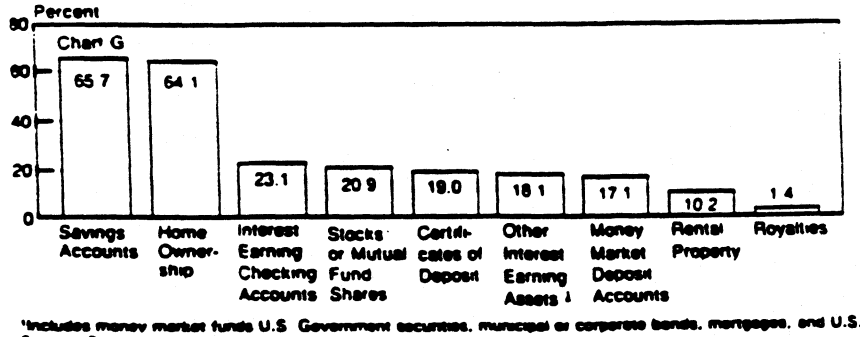


Asset Ownership

Chart G shows the percent of households reporting ownership of asset types for the period June through December, 1983. Not surprisingly the most common forms of asset ownership are savings accounts and home ownership. Over twice the percentage of households own savings accounts and homes than any other asset on the chart. The least common forms of asset ownership among households are rental property and royalties.

Households, By Ownership Of Asset Types

(June - December 1983)



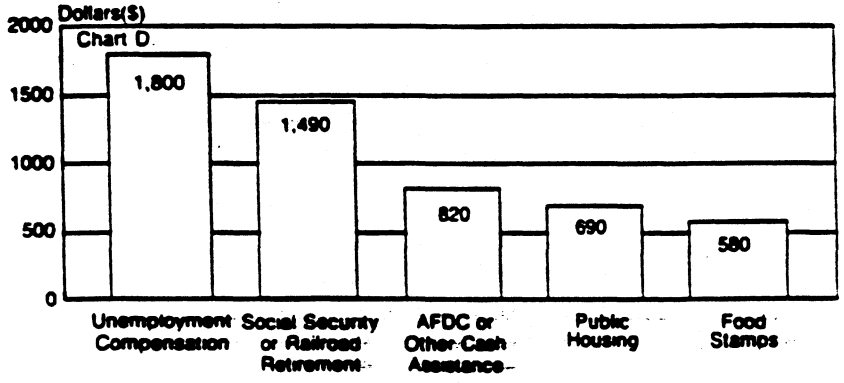
[1]Includes money market funds U.S. Government securities, municipal or corporate bonds, mortgages, and U.S. Savings Bonds

Program Participation

There are a variety of government programs in which people participate, such as Social Security or Railroad Retirement, unemployment compensation, food stamps, Aid to Families with Dependent Children (AFDC) and public housing. One of the main goals of the SIPP is to provide better information on such program participation. For example, participation in the programs just mentioned can be viewed in relationship with monthly income. As should be expected, there are differences in the average monthly income received by households participating in these programs. The households receiving unemployment compensation show an average monthly income of $1,800. Those households receiving Social Security or Railroad Retirement had an average monthly income of $1,490. Those households receiving AFDC or other cash assistance show an average monthly income of $820. The residents of public housing had an average monthly income of $690. Households receiving food stamps received an average monthly income of $580.

Households, by Average Monthly Income and Selected Program Participation

(Third Quarter 1983)



Labor Force Activity

The data from SIPP provide some insights into the relationship between the income of a household and the labor force status of its members. Out of 83.1 million total nonfarm households, 53.3 million contain one or more members holding a job in the third quarter of 1983 (and no one looking for work or on layoff). These households had an average monthly income of $2,680. There were 11.5 million households with one or more members looking for work or on layoff. These households had an average monthly income of $1,650. The third group of households, 18.3 million with no members with labor force activity, had an average monthly income of $1,000.

APPENDIX K

SIPP Quarterly Report Tables

Each report contains text tables and detailed tables. The text tables are designed to focus on a different topic in each report. For example, the fourth quarter 1983 report featured data on health insurance coverage. The detailed tables, on the other hand, contain the same type of information each time. Following is a list of the standard detailed tables that appear in each report:

1. Selected characteristics of persons, by monthly household cash income: monthly average.
2. Selected characteristics of persons, by mean monthly household cash income and program participation status: monthly average.
3. Persons, by labor force activity status, age, sex, mean monthly household cash income, and program participation status: monthly average.
4. Selected characteristics of persons 16 years old and over, by monthly earnings: monthly average.
5. Households, by labor force activity status of members, mean monthly household cash income, and program participation status: monthly average.
6. Selected characteristics of households, by monthly household cash income: monthly average.
7. Selected characteristics of households, by mean monthly household cash income and program participation status: monthly average.
8. Households, by receipt of selected sources of noncash benefits and money transfer payments: monthly average.
9. Households, by monthly household cash income and selected sources of money income and noncash benefits: monthly average.

APPENDIX J

Schedule for Releasing SIPP Quarterly Reports and Core Public-Use Microdata Files

NOTE: As of December 1985, quarterly reports from the third quarter of 1983 through the fourth quarter of 1984 are available as well as public-use files for core data from the first 5 waves of interviewing. To obtain public-use files, write to: Data User Services Division, Bureau of the Census, Washington, DC 20233, or call (301) 763-4100. Copies of released reports are for sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402. (The order title is: Current Population Reports, Series P-70, Economic Characteristics of Households in the United States.)

First Quarter 1985 Report.....1/86

Second Quarter 1985 Report and Wave 6 (1984) Public-Use File.....3/86

Third Quarter 1985 Report and Wave 7 (1984) Public-Use File.....6/86

Fourth Quarter 1985 Report and Wave 8 (1984) Public-Use File.....9/86

First Quarter 1986 Report and Wave 9 (1984) Public-Use File.....1/87

Herriot, R.A. and D. Kasprzyk (1984). The Survey of Income and Program Participation, SIPP Working Paper Series No. 3405. Washington, D.C.: U.S. Bureau of the Census.

Kasprzyk, D. and D. Frankel (1985). Survey of Income and Program Participation and Related Longitudinal Surveys: 1984: Selected papers given at the 1984 Annual Meeting of the American Statistical Association. Washington, D.C.: U.S. Bureau of the Census.

McMillen, D. and R.A. Herriot (December 1985). "Toward a Longitudinal Definition of Households," Journal of Economic and Social Measurement, Vol. 13, No. 3-4, New York: Elsevier Science Publishing Co., Inc.

Short, K. (1985). The Survey of Income and Program Participation: Uses and Applications, SIPP Working Paper Series No. 8501. Washington, D.C.: U.S. Bureau of the Census.

REFERENCES

References Cited in the Text

Coder, J. and A. Feldman (1984). "Early Indications of Item Nonresponse on SIPP," Proceedings of the Survey Research Methods Section, American Statistical Association.

David, Martin H., Ed. (1983). Technical, Conceptual, and Administrative Lessons of the Income Survey Development Program (ISDP). New York: Social Science Research Council. (Out of Print)

Frankel, Delma (1985). Summary of the Content of the 1984 Panel of the Survey of Income and Program Participation, SIPP Working Paper Series No. 8504. Washington, D.C.: U.S. Bureau of the Census.

Haber, Sheldon E. (1985). Applications of a Matched File Linking the Bureau of the Census Survey of Income and Program Participation and Economic Data, SIPP Working Paper Series No. 8502. Washington, D.C.: U.S. Bureau of the Census.

Jean, A. and E. McArthur (1984). "Some Data Collection Issues for Panel Surveys with Application to the Survey of Income and Program Participation," Proceedings of the Survey Research Methods Section, American Statistical Association.

Kalton, Graham and James Lepkowski (December 1985). "Following Rules in the Survey of Income and Program Participation," Journal of Economic and Social Measurement, Vol. 13, No. 3-4, New York: Elsevier Science Publishing Co., Inc.

Lamas, E.J. and J.M. McNeil (1984). "The Measurement of Household Wealth in SIPP." Proceedings of the Survey Research Methods Section, American Statistical Association.

McMillen, D., and D. Kasprzyk (1985). "Item Nonresponse in SIPP." To be published in the Proceedings of the Survey Research Methods Section, American Statistical Association.

Ryscavage, Paul M. and John E. Bregger (September 1985). "New Household Survey and the CPS: Labor Force Differences," Monthly Labor Review, Vol. 108, No. 9. Washington, D.C.: Bureau of Labor Statistics.

Ycas, Martynas and Charles Lininger (November 1981). "The Income Survey Development Program: Design Features and Initial Findings," Social Security Bulletin, Vol. 44, No. 11. Baltimore, MD: Social Security Administration.

References Not Cited in the Text

Frankel, D. (1985). The Survey of Income and Program Participation: 1985. Selected papers given at the 1985 annual meetings of the American Statistical Association. Washington, D.C.: U.S. Bureau of the Census.