THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

AN OVERVIEW OF THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

No. 01

Dawn Nelson
David McMillen
Daniel Kasprzyk

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

Survey of Income and Program Participation

Working Paper Series

AN OVERVIEW OF THE SURVEY OF
INCOME AND PROGRAM PARTICIPATION

by

Dawn Nelson, David McMillen,
Daniel Kasprzyk

No. 8401 (1)

U.S. Department of Commerce
Malcolm Baldrige, Secretary
Clarence J. Brown, Deputy Secretary
Sidney Jones, Under Secretary for
Economic Affairs

BUREAU OF THE CENSUS
John G. Keane,
Director

AN OVERVIEW OF THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

Introduction and Background

The Survey of Income and Program Participation (SIPP) is expected to become a major source of information on the economic situation of persons and families in the United States. The survey will provide data for a better understanding of the income distribution in this country. In addition, the data will be used to study Federal and State transfer and service programs, to estimate future program costs and coverage, and to assess the effects of proposed changes in program eligibility rules or benefit levels. The data will also provide information for debating policy issues such as national pension and retirement plans, tax reform, social security funding, and health care reform.

Currently, the main source of data on income and program participation is the March Income Supplement of the Current Population Survey (CPS). Although the CPS March Supplement provides the best income data now available on a yearly basis, it is not designed to meet certain critical data needs relating to the measurement of cash and noncash income. This is in large part due to the principal purpose of the CPS, which is to obtain information on employment and unemployment, and the consequent constraints on space and time to ask additional questions. As a result of these constraints, three basic shortcomings are evident: (1) underreporting of information that is difficult to recall in detail over a long reference period, such as property income, irregular income, and program participation; (2) misclassification of participation in major income security programs; and (3) insufficient information necessary to fully analyze program participation and eligibility.

Awareness of these limitations led to support for a "new income survey" in the early 1970's. The Department of Health, Education, and Welfare responded

to the growing demand for improved data and established the Income Survey Development Program (ISDP) in 1975 to plan the development of a new survey. During the next 5 years, extensive research was undertaken on ways to improve the measurement, collection, and processing of income, program, and wealth data. Much of the work centered around four experimental field tests that were conducted in collaboration with the Bureau of the Census 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 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 8,200 households and provided a vehicle for feasibility tests and controlled experiments of alternative design features. Although used primarily for methodological purposes, it was sufficiently large to provide reliable national estimates of many characteristics of interest to analysts. Public-use microdata files and documentation of the 1979 Research Panel are available through the National Technical Information Service.1/ Attachment A provides a more complete description of the ISDP and of the initial findings from the 1979 Research Panel. Additional information on the ISDP, including a bibliography of ISDP materials, can be found in The Technical, Conceptual, and Administrative Lessons of the Income Survey Development Program (Martin David, ed. New York, Social Science Research Council, 1983).

SIPP Data Uses

Analyses of the ISDP tests, experiments, and data have been used in planning

---

1/ 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.

SIPP to help meet the need for several types of data currently unavailable or not available with sufficient accuracy or frequency, including:

- o Income data for each month in the calendar year.
- o Income data based on a wider variety of cash and noncash sources.
- o Data to estimate after-tax income distribution.
- o Annual household income data adjusted for monthly changes in household composition.
- o Monthly data on most government income transfer and service programs.
- o Detailed data on assets and liabilities.
- o Information for determining program eligibility and defining program recipient units.

Federal agencies will use SIPP data to address policy and administrative issues related to a broad spectrum of income security and social welfare programs. For example, how many persons and families eligible for benefits are not receiving them; if eligible persons were to receive benefits, how would Federal programs need to be adjusted? Or, how would changes in eligibility rules or benefit levels affect recipients and program costs? The Department of Agriculture is interested in studying how rules on counting assets affect eligibility for the food stamp and school lunch programs. Similarly, the Department of Housing and Urban Development, wants to use the data to study federally subsidized public housing and mortgage program issues. Broader issues affecting all programs will also be examined using SIPP data; for example, the Treasury Department will use the data in microsimulation modelling to observe the results of alternative taxing arrangements. In addition, a number of academic and business organizations have shown considerable interest in using SIPP data for a variety of research projects. The following sections of this paper describe various aspects of SIPP that are expected to contribute to achieving the goals and meeting the needs expressed by the many diverse users. The sections cover: 1) the sample design;

2) survey content; 3) operational features, including field procedures and processing plans; and 4) the Census Bureau's preliminary plans for data products.

Sample Design

SIPP started in October 1983 as an ongoing survey program of the Bureau of the Census with a sample of approximately 25,000 "designated" housing units selected to represent the noninstitutional population of the United States. The "assigned" sample size will be somewhat smaller (about 20,000 households) because some of the selected units will be unoccupied, demolished, converted for nonresidential use, or occupied by persons not eligible for interview such as persons maintaining their usual residence elsewhere. Attachment B provides a brief description of the sample for this panel, referred to as the "1984 Panel." Attachment C describes the operations of this panel which are somewhat different from subsequent panels due to factors associated with starting such a system.

Each assigned household will be interviewed once every 4 months for 2 1/2 years to produce sufficient data for longitudinal analyses while providing a relatively short recall period for reporting monthly income. The reference period is the 4 months preceding the interview; for example, in October, the reference period is June through September. When the household is interviewed again in February, it is October through January. This interviewing plan will result in eight interviews per household. A shorter reference period, such as the 3-month period tested in the ISDP surveys, was rejected because it would have been much more expensive to interview the same size sample for 2 1/2 years, and more frequent visits or a longer panel period might substantially increase the nonresponse rate which would introduce biases and reduce sampling efficiency.

In January 1985 and every January thereafter, a new, slightly smaller panel will be introduced. (Attachment D outlines the plan for implementing sample panels

in 1984 through 1987.) This design will allow cross-sectional estimates to be produced from a combined sample of approximately 35,000 households. This overlapping panel design will also enhance the estimates of change, particularly year-to-year change. 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 group of households. This design will give a more precise estimate of change than a similar comparison involving two different groups of households in which greater sampling variability obscures the actual change.

Finally, to facilitate field operations, each sample panel is divided into four approximately equal subsamples, called rotation groups; one rotation group will be interviewed in a given month. Thus, one cycle or "wave" of interviewing takes 4 consecutive months. This design creates manageable interviewing and processing workloads each month instead of one large workload every four months; however, it results in each rotation group using a different reference period. Figure 1 provides an illustration of the relationship between waves, rotation groups, interview months, and reference periods after the sample becomes fully operational in 1985.

Survey Content

There are four key questionnaire components for collecting SIPP data: 1) the control card, 2) the core set of questions repeated on each wave questionnaire, 3) the fixed topical modules assigned to specific waves, and 4) the variable topical modules to be added from time to time.2/ In addition, the survey questionnaire content may be supplemented with administrative record data which are difficult to obtain in a survey such as lifetime earnings and program benefit histories. To

---

2/ To obtain copies of the control card and/or the questionnaire for a particular wave of interviewing, write to: Income Surveys Branch, Demographic Surveys Division, Room 3339, Bureau of the Census, Washington, DC 20233, or call (301) 763-2063.

Figure 1. SIPP Interview Months and Reference Periods

Wave	Rotation Group	Interview Month	Reference Periods																			
							1st Qtr Date			2nd Qtr Date			3rd Qtr Date			4th Qtr Date						
1	1	JAN	SEP	OCT	NOV	DEC																
	2	FEB		OCT	NOV	DEC	JAN															
	3	MAR			NOV	DEC	JAN	FEB														
	4	APR				DEC	JAN	FEB	MAR													
2	1	MAY					JAN	FEB	MAR	APR												
	2	JUN						FEB	MAR	APR	MAY											
	3	JUL							MAR	APR	MAY	JUN										
	4	AUG								APR	MAY	JUN	JUL									
3	1	SEP									MAY	JUN	JUL	AUG								
	2	OCT										JUN	JUL	AUG	SEP							
	3	NOV											JUL	AUG	SEP	OCT						
	4	DEC												AUG	SEP	OCT	NOV					
4	1	JAN													SEP	OCT	NOV	DEC				
	2	FEB														OCT	NOV	DEC	JAN			
	3	MAR															NOV	DEC	JAN	FEB		
	4	APR																DEC	JAN	FEB	MAR	
	1	MAY																	JAN	FEB	MAR	APR

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 attachment E).

Control Card. The control card is used to obtain and maintain information on the basic characteristics associated with households and persons, and to record information for operational control purposes. Characteristics recorded on the control card by the interviewer include the age, race, ethnic origin, sex, marital status, and educational level of each member of the household, as well as information on the housing unit and the relationships among members of the household. A household respondent provides this information, which is updated, as appropriate, at each interview. The control card is also used to keep track of when persons enter and leave the household. Thus, a monthly record of changes in household composition will be available for use in generating point-in-time estimates, longitudinal estimates, and analyses of change over time. Space is also available to record information that will improve our ability to follow persons who move during the survey. In addition, after each visit, data on employment, income, etc., are transcribed from the core questionnaire to the control card so the data can be used in the next interview to assist in monitoring changes in employment status and income recipiency.

Core Questions. Questionnaire items included in the "core" of SIPP are directed at assessing the economic situation of persons in the United States and are repeated in each wave of interviewing. These questions mainly cover labor force participation and amounts and types of income received, including transfer payments and noncash benefits from various programs for each month of the reference period. A few questions on other topics such as coverage by private health insurance plans are also included in the core.

At each interview, the core questions are used to determine a person's labor force status during the past 4 months. Persons who were employed any time during the reference period are asked to report on jobs held or businesses/farms owned, number of hours and weeks worked, earnings, and any weeks without a job or business. Questions for those who were not working some or all of the period are designed to determine if they should be classified as unemployed, discouraged workers, or not in the labor force.

In addition to questions about income associated with labor force activity, such as farm and nonfarm wages and salaries and self-employment income, the core questions cover nearly 50 other types of income. For example, questions are asked to prompt reporting of government transfer payments from retirement, disability, and unemployment benefit and welfare programs such as food stamps. Information on the receipt of noncash benefits from programs such as Medicare and Medicaid is also obtained. Other income questions in the core relate to private transfers such as pensions from employers, alimony, and child support; however, some private noncash benefits, such as stock options, are not covered. There are also questions on ownership of assets that produce income, such as interest, dividends, rent, and royalties, and on miscellaneous sources of income, such as estates. A complete list of the types of income covered in SIPP and preliminary definitions are given in attachment F.

Fixed Topical Modules. The content of SIPP will be expanded in Waves 2 through 8 for each panel after 1984 by adding questions on topics not covered in the core section. These fixed topical modules are assigned to a particular wave and may be repeated in a later wave if more than one observation is needed. The topics covered in these modules do not require repeated measurement during the year and may use a longer reference period than that for the core questions. For example,

a wealth module will be administered twice in each panel, in waves one year apart, to collect detailed data on personal and household assets and liabilities (see attachment G for a list of the data items). These data will allow an examination of economic well-being beyond that which can be observed through the study of income. An annual "round-up" module will be administered in the waves at the end of the first and second years of interviewing to obtain wage and salary data from W-2 forms, tax-filing status, and estimates of annual self-employment and property income, and taxes paid. This module will also gather information on employer-provided benefits and educational enrollment. Two other topical modules will be administered in only one wave of a panel to obtain: 1) marital history, fertility, and migration data; and 2) education, health and disability, and work history data. A schedule identifying the waves in which these modules will be administered is provided in attachment H.

Variable Topical Modules. Variable topical modules make up the final component of the SIPP questionnaire. These modules will include supplemental questions designed by or for other Federal agencies and will be added to one or more waves of interviewing. The core questions and the fixed topical module questions have been designed to allow adequate time for additional questions during some of the interviews. This will provide the flexibility to be responsive to other agencies' data requirements for program planning and policy analysis. The variable topical modules have already been identified for the first SIPP panel (see attachment H); other modules for future panels will be planned as issues emerge.

Operational Features

Field Procedures. Data collection operations will be managed through the Census Bureau's 12 permanent regional offices. Interviewers assigned to these offices will conduct one personal visit interview with each sampled

household every 4 months. At the time of the interviewer's visit, each person 15 years old or older who is present will be asked to provide information about himself/herself; a proxy respondent will be asked to provide information for those who are not available. Telephone interviewing will be permitted only to obtain missing information, to interview persons who will not or cannot participate otherwise, or to interview persons who have moved more than 100 miles from a SIPP PSU.

The longitudinal design of SIPP dictates that all persons 15 years old or older in a household at the time of the first interview remain a part of the survey even if they move to a new address during the next 2 1/2 years. This design feature requires several special field procedures which are described in attachment I. For example, the interviewer must obtain the name, address, and telephone number of a person who could supply the new address should the entire household move. For cost reasons, personal visit interviews are only conducted at new addresses that are within or reasonably close to a SIPP sampling area. The geographic area defined by those rules for personal visit interviewing at a new address contains over 95 percent of the population. Telephone interviews are encouraged for persons living beyond the personal visit follow-up area.

In addition to an interest in the experience of individuals over time, this survey is also designed to capture the household experiences of those persons since the household is a primary unit of economic analysis. Consequently, SIPP field procedures call for interviewing any adult living with a person included in the first wave of interviewing. For similar reasons, changes in household composition caused by persons who join or leave the household are recorded at each interview.

To facilitate keeping track of respondents in the field as well as the linking of data across waves, each person is assigned a unique, unchanging 14-digit identification number at the time he/she enters the sample. Once assigned, the person ID is never changed, except for rare instances of the merging of two or more sample households into one household.

In addition to unique, unchanging person ID's, a two-digit address ID code is also assigned. The address ID code is used along with nine digits of the person ID to identify separate households in any given wave. The identification scheme provides a means of linking information about an individual across time and identifies which household each person is a member of at any wave in the panel. Additional processing using control card information, such as "date entered the household" and "date left the household" is then done to group persons into households for each survey month; thus monthly household determination occurs during processing and does not affect operations, except for the collection of entry and exit dates. Thus, it is possible to link data from all persons ever associated with a given household throughout the 2 1/2-year duration of a panel. This will facilitate 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 is provided in attachment J. A master list of identification numbers will be used by the regional offices in monitoring the status of interviewing each month after Wave 1. The regional offices will be responsible for ensuring that there is a completed questionnaire (or the reason for noninterview) for each number on the list representing all the persons assigned for interview in a month. The list will be updated regularly to account for persons who are added or deleted from the sample.

Processing Plans. The questionnaires and control cards received by the regional offices will be processed on a monthly basis. This processing includes: 1) data keying; 2) the creation of person, family, and household records from the data; 3) the addition of industry and occupation codes, and geographic codes to the appropriate records; 4) several types of editing (see attachment K); 5) imputation for missing data (see attachment L); and 6) weighting (see attachment M). Several means of monitoring the processing will be built into the system, including a series of regional office progress reports and computer-generated reports designed to show when production is behind schedule as well as other problems.

One advantage of a monthly processing system is the ability to improve the quality of the data by resolving problems on a timely basis. For example, errors discovered during the preedit (the first extensive computer edit) can be corrected by contacting the interviewer or respondent, and interviewers can be informed about recurrent problems to improve their future performance. Monthly processing will also accelerate the production of quarterly cross-sectional reports. If the questionnaires were processed on a wave basis, the processing would take longer because the data referring to any one calendar quarter are collected in more than one wave. This is a result of the SIPP rotation group pattern and reference period scheme (see figure 1.)

Preliminary Plans for Data Products

A number of publications and public-use data files will be generated from the information collected in SIPP. Both publications and data files are differentiated by whether they are cross-sectional or longitudinal. Two types of cross-sectional reports are planned by the Census Bureau: 1) a set of quarterly and annual reports that will focus on core information; and 2) a set of periodic or single-time reports that will use the detailed sociodemographic data from the topical modules.

Quarterly Reports. These reports will provide aggregate values for a variety of labor force, income, and household composition measures based on monthly averages. The annual reports will be similar in content, but will show values averaged across 12 months rather than 3 months. The first quarterly report will be issued in fall 1984 and will contain income and labor force data referring to the third quarter of 1983. It is anticipated that the lag between collection and publication will eventually be reduced to the point where quarterly reports will be released in the first month of that quarter 1 year later, or 6 months after the last data are collected for that quarter. Further reductions in the lag are impeded by the monthly interviewing scheme which requires sample data from 6 months of interviewing to produce a quarterly estimate (see figure 1). A tentative list of a few table titles suggested for the quarterly reports is given in attachment N. Attachment O displays examples of the types of tables that might be prepared for this report series.

Periodic or Single-Time Reports. Detailed data from the topical modules will be used to examine issues related 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.

Cross-Sectional Data Files. SIPP cross-sectional data files will be issued on a wave-by-wave basis approximately one month after the published data are released. Each file will include person, family, and monthly household information collected in a given wave of the survey. During the reformatting process, some of the collected data will be excluded or summarized to create a standard format across waves. In addition, data that might disclose the identity of a person will be

excluded or recoded in accordance with standard Census Bureau confidentiality restrictions. These files will be fully edited, imputed, and weighted in a manner consistent with their use for cross-sectional analyses. They will contain sufficient information for users who want to create quarterly files, or files covering some other time period. A longitudinal file can be created by matching two or more wave files using the unique person ID number; however, it will be the user's responsibility to develop longitudinal edits, allocation values, and weights for these merged files.

Longitudinal Reports. Plans for the publication of longitudinal reports from the SIPP are now under development. Six types of reports are being considered: 1) economic profile; 2) comparative profile; 3) transition; 4) multiple; 5) longitudinal family and unrelated individual; and 6) special event.

The time dimension is an issue arising in the discussion of these kinds of reports; for example, it is possible to report year-to-year, quarter-to-quarter, and even month-to-month changes in the SIPP. Obviously, reports using SIPP longitudinal data can be far-reaching and extensive. The six types of reports under consideration are described below:

1) Economic Profile Reports. These publications present data on income and program participation for persons. These data are yearly aggregates of monthly data as reported in the SIPP core questionnaire. The tables, similar to those published in the quarterly cross-sectional reports, provide an economic profile of individuals on a calendar-year basis. Unlike the annual cross-sectional reports which provide monthly values averaged over a 12-month period, these reports provide the aggregation of monthly income values during the year.

In addition, all information on households and families will be reported as characteristics of individuals. Household and family characteristics as well as aggregated data across time are assigned to individuals. For example, this will provide measures such as the number of persons living in households with an average monthly household income of $2,000 to $3,000. These reports would not provide data on longitudinal households and families.

2) Comparative Profile Reports. These reports extend the economic profile reports to allow the comparison of yearly aggregates across time. This would allow, for example, a detailed comparison of the economic profiles of persons for 1984 and 1985.

3) Transition Reports. These reports could focus on the change in status between two points in time. Changes in annual, quarterly, or monthly status, or even changes between the beginning and end of the year could be reported. Thus, a typical table could provide the number and percent of persons with X dollars of annual income in one year and Y dollars of annual income in the next year. Another way of conceptualizing the information provided in these reports is that they provide the information necessary to develop a set of social accounts. That is, the population is distributed across a set of statuses in January and that distribution is somewhat different at another point in time. The transitions reported in these reports describe how the population moved from the first distribution to the second.

4) Multiple Transition Report. This type of report is characterized by an extension from single to multiple transitions. These tables indicate the economic and demographic characteristics of persons who changed status more than one time; for example, from being employed to unemployed to employed again, or from receiving a program benefit to not receiving and then to receiving again. In addition to reporting on these patterns of activity, these reports describe

the number of spells or experiences in a given state as well as the durations of those spells. An alternative characterization of these reports is that they describe and summarize the monthly and/or quarterly observations. The patterns that exist across time provide the description of activity, and the data on spells and durations serve as summary statistics of that activity.

5) Longitudinal Family and Unrelated Individual Reports. These reports describe the characteristics of the longitudinal family units defined in the SIPP, as well as the unrelated individuals who complete the sample universe. Detailed characteristics, such as average family size and the presence of persons in special age groups such as children or the elderly, comprise one dimension of these reports. In addition, annual aggregates for these family units can be presented. Families that did not exist for the full longitudinal period can be included by time-weighting their values or including them in a separate table. Finally, statistics on unrelated individuals can be tabulated to conform wherever possible to the tabulation for families.

6) Special Event Reports. This series of reports will include data that are collected which relate to a unique event or current data tabulated with a particular focus. Thus, the reports could include multi-wave data collected in the core combined with topical module data such as wealth or work history. The reports could also focus on special populations, such as the elderly or food stamp recipients, or on events which change the economic situation of an individual; for example, marriage, divorce, separation, the birth of a child, a return to school, or a move to a new address are events which are "special" in the sense that the economic situation of the person, both prior to and after the event, is of interest. These reports differ from other periodic reports only in that they report longitudinal rather than cross-sectional data.

Longitudinal Microdata Files. The basic longitudinal data file being planned for SIPP is a calendar-year file; that is, the data items on the file will be monthly observations for the 12 months of a calendar year for each person in the sample. Thus, the first of this series would be a calendar year (CY) 1984 file from the 1984 Panel. The second of the series would be a CY85 file produced in two parts--a CY85 file from the 1985 Panel, and a CY85 file from the 1984 Panel. These files will be prepared in such a manner that each can be used separately or, with a simple adjustment to the weights, together. Monthly data from what is considered to be the "core" of the questionnaire will be contained on these files.

The CY84 file will be produced from interviews in Waves 2 through 5 of the 1984 Panel. The file is likely to be a person-level file with monthly amounts of income and program-related data. These data will not identify longitudinal units, but rather provide the monthly information necessary for the user to construct various types of longitudinal units. If, however, the current exploration of alternative methods of defining longitudinal families leads to an acceptable definition, longitudinal family indicators could be added to this file. In addition to the calendar-year data files, the Population Division will develop, for internal use, a longitudinal file of interviews conducted in Waves 1, 2, and 3. With appropriate editing for confidentiality, this file may be released as a research file; however, it is likely that this file will not be fully edited and imputed.

Finally, a longitudinal file will be developed separately for each panel. That is to say, a 1984 Panel file will be created containing Waves 1 to 9 interviews as well as any topical modules. Similarly, a 1985 full Panel file will be developed with the Waves 1 to 8 core interview data and any topical module data gathered.

ATTACHMENT A

The Income Survey Development Program:

Design Features and Initial Findings

Reprinted from the **Social Security Bulletin,** November 1981
U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

Social Security Administration
SSA Publication No. 13-11700

The Income Survey Development Program: Design Features and Initial Findings

by Martynas A. Yčas and Charles A. Lininger *

This article summarizes the principal design features of the Income Survey Development Program (ISDP) and presents initial findings of its 1979 Research Panel. The ISDP was designed to meet the need for improved data—particularly information on cash and in-kind income, assets and debts, tax liabilities, and participation in the major income security programs. The ISDP examined many technical and operational problems that were difficult to satisfy with existing surveys and administrative record systems. The ISDP field tests indicated that several experimental features were successful. These include use of more frequent interviews, a sample that is followed over a period of more than a year, a flexible questionnaire structure that permits insertion of questions on emerging policy issues, and procedures to maximize the linking of survey data with information in administrative records. Initial findings from the 1979 Research Panel indicate the number of persons receiving benefits from more than one major transfer program and compare the assets of food stamp recipients with those of eligible nonparticipants and ineligible persons.

A large share of the budgets of the Department of Health and Human Services (HHS) and other agencies is spent in providing cash payments and services to individuals and families on the basis of need and entitlement. The cost, impact, and fairness of these programs are of executive and legislative concern. Yet the information now available has some serious deficiencies. Many persons and families receive benefits from more than one program—for example, Social Security, Supplemental Security Income, food stamps, Aid to Families with Dependent Children, and Medicaid—and this overlapping may lead to great disparities in the amount of assistance received by some portions of the needy population. To a varying and sometimes considerable extent, persons who are eligible for benefits from a particular program do not apply for them; others pyramid benefits from different sources and the cumulative effect may be greater than program officials intended. As a result, some target groups may not get the assistance intended for them and others may receive more than was anticipated. Needs may therefore be imperfectly or inefficiently met. The existing poverty measures are based on annual cash income and do not take account of assets, benefits in kind, and fluctuations in income during a year. When changes are considered in eligibility rules and benefit schedules, these limitations make it difficult to estimate the cost to the various levels of government and to answer policy concerns as to who will win and lose.

The main regular source of information on household and personal income in the United States is the March supplement to the Current Population Survey (CPS). The CPS is designed to provide timely and accurate statistics on employment and unemployment. To achieve these objectives at a reasonable cost, it relies on brief interviews, usually by telephone, with whoever is living at a sample address. This procedure provides enough time for labor-force questions and a limited number of questions on supplementary topics—for example, in March of each year, questions about annual cash income and receipt of a few major in-kind benefits. These necessarily brief questions succeed in measuring regular and easily recalled types of income, but fail (according to estimates based on administrative records) to account for between 25 percent and 50 percent of income from public assistance and other government transfer payments except Social Security, about 25 percent of income from public and private pensions, and more than

* Income Survey Development Program Staff, Office of Research and Statistics, Office of Policy, Social Security Administration.

half of property income. Assets, taxes, and expenses used in the computation of eligibility and benefit amounts are not measured at all, nor are monthly and quarterly fluctations in income. Questions in the March supplement on annual income over the previous calendar year are asked only of household members present at the time of interview. They omit the income of persons who have died, have been institutionalized, have left the county, or have left the household because of marriage, divorce or separation, or for other reasons. No account is taken of changes in household membership that may have had a major impact on the financial well-being of those who resided together during the reference year.

These problems may well be intrinsic to the design of the CPS. The changes and additions that would be required to improve its income and program data and to add tax and asset data would compromise the survey's continued success in carrying out its major function.

Analysts have therefore relied heavily on special surveys with a single interview per household and a primary focus on income-related topics. These surveys have included the Survey of Economic Opportunity in 1967 and the Survey of Income and Education in 1976, both of which attempted to collect more detailed income data, and the Survey of Financial Characteristics of Consumers, the most recent (1963) comprehensive measure of the distribution of asset holdings. These one-time efforts did not entirely succeed in collecting the desired data, and inevitably their results become less useful as they become more dated.

The awareness of these limitations in other surveys led in 1970 to proposals for a new income survey that would do a better job of capturing information on income, taxes, assets, and program participation. In 1975, a formal effort was launched to see how such a data-gathering mechanism might be structured. This effort, called the Income Survey Development Program (ISDP) was directed by the Office of the Assistant Secretary for Planning and Evaluation in HHS and was carried out jointly by the Bureau of the Census, which participated in planning and carried out the field work, and the Social Security Administration (SSA), which administers the major cash income security programs.

Design Features of ISDP

The ISDP was designed to test new methods of collecting more complete and accurate income, asset, and program participation data. Between 1977 and 1980, the ISDP carried out four major field tests of increasing complexity. A Site Research Test was carried out in five cities in 1977-78. A nationwide 1978 Research Panel was conducted in 1978-79, and a larger 1979 Research Panel was conducted in 1979-80. The Special Frames Study examined subpopulations drawn from six diverse administrative record systems in five States during the summer of 1980. The primary purpose of these surveys was methodological; they combined feasibility tests with controlled experimental comparisons of alternative design features. The large, specially stratified national sample of the 1979 Research Panel, however, was sufficient to provide nationally reliable estimates of many characteristics of interest to analysts.

The ISDP was intended to overcome three deficiencies of other surveys. First was the need for information about resources that are difficult to remember or to account for in ways useful to analysts. For example, it is difficult to obtain monthly detail on income received during the year, especially if it has changed significantly from month to month because of variations in benefit receipt or because of casual or intermittent earnings.

Second was the need for a way to collect the large amount of information required to analyze program participation and eligibility. Answering the required number of questions places a large burden on survey respondents. Detail on some 50 kinds of income received over different periods of time—weekly, monthly, quarterly—must be collected so that it can be made to correspond to program rules. It is also virtually essential to have detailed data on labor-force participation, expenditures that may offset income in calculating eligibility (medical, educational, child care, and work-related costs), household composition, health status, asset holdings (both as sources of income and as determinants of eligibility for programs with asset screens), taxes (to calculate disposable income), and the receipt of various goods and services that might be considered as income in kind.

Third was the need for flexibility. Changes in the economy and in political priorities bring new issues into prominence and change the terms in which old issues are discussed. The structure developed for the ISDP had to focus on groups and issues of emerging interest within a relatively short lead time.

The basic approach used was a panel design. Persons at sample addresses were contacted early in the calendar year and asked about their income and other characteristics for the previous few months. They were then recontacted at regular intervals, usually every 3 months. Persons who moved, individually or in family groups, were interviewed at their new addresses in order to maintain the integrity of the sample and obtain full detail on changes in income and household composition over an entire year. This procedure was continued until the spring of the following year, and was concluded with a set of questions on taxes. As a result, a detailed record was built for each person for the entire calendar year. This procedure minimized the need for respondents to remember income receipt for more than 3 months; it also minimized the number of questions that had to be answered in each interview.

Because less time was required to update the basic or core information after the initial interview, time was also available in later "waves" of interviewing to ask additional questions on topics that were either stable enough not to require quarterly updating—marital history and pension coverage, for example—or were emerging issues of one-time interest—for example, emergency energy assistance received from the special program during the winter of 1979-80. This design enabled the survey staff to devise a set of core questions on income and other eligibility determinants well in advance, ensuring timely processing and rapid turnaround, while leaving interview time for the addition of questions on new policy issues with a shorter lead time.

Table 1 shows the timing and content of the 1979 Research Panel. It shows the variety of data that can be accumulated on families and individuals under the panel approach. It also displays the 1979 Panel data that are now being made available to interested users.

Another approach taken to solve the problems of burden and accuracy was to maximize the linkage of survey responses with data in administrative record systems. Some of these systems, such as the summary earnings records maintained by SSA, contain detailed information extending over many years that would be impossible to collect accurately in personal interviews.

It was therefore necessary to test whether accurate Social Security numbers (SSN's) could be collected, since this number is the identifier in most general use. Despite early concern that collection would be sensitive, less than 3 percent of the persons in the 1979 sample refused to provide this information. Validation of reported numbers and followup questions for nonmatching cases—an advantage permitted by the panel design—resulted in valid SSN's for more than 95 percent of the sample cases.

The ISDP staff also explored the feasibility of drawing samples from beneficiary records. Such specially selected samples can provide enough cases to examine relatively small program populations of policy interest if this kind of analysis is desired from an operational survey program. Further, the answers of respondents in the samples of the research panels were matched to administrative records in order to validate the responses and to assess the questionnaire. At various times the ISDP surveys included samples of persons receiving Supplemental Security Income (SSI); Aid to Families with Dependent Children (AFDC); Old-Age, Survivors, and Disability Insurance (OASDI); Basic Educational Opportunity Grants, now Pell Grants; unemployment compensation; workers' compensation; veterans' benefits; and also included certain categories of tax filers.

During the development period the ISDP staff devised and tested a number of specific variations in survey methods. The "standard" questionnaire format used in most of the interviews was person-based, so that a separate form was filled out for each adult in a sample household, and extensive use was made of skip patterns and check items to ensure that a probing, highly specific set of questions appropriate to each person's status and probable income sources would be asked. This procedure, though effective, requires a relatively lengthy interview and was adopted after two alternatives were tested in the Site Research Test in 1977-78. The long form was used with half the sample, and an alternative short form that asked everyone a direct, nonprobing set of questions on all income types of interest was used with the other half. As expected, the latter approach did shorten the length of interviews somewhat, but it also reduced response accuracy and was often judged by the interviewee to be boring. Another alternative, modeled on the revised CPS income supplement, was tested with the 1979 Research Panel. This "B" form screened the entire household for receipt of various types of income by asking a single respondent about other members of the household. Again, interview time decreased slightly at the expense of reporting accuracy.

Another parameter that was varied in ISDP tests was the choice of respondent. During most interviews, proxy responses were accepted from other household members when convenient, although it was expected that these re-

Table 1.—Staggered interview design of 1979 Research Panel[1]

Content	First rotation group	Second rotation group	Third rotation group
First wave: Household composition; labor-force participation; income profile; attitudes	February 1979	March 1979	April 1979
Second wave: Household composition, labor-force participation, and income update; assets; shelter, medical, and work-related costs; attitudes	May 1979	June 1979	July 1979
Third wave: Household composition, labor-force participation, and income update; work and marital history; educational attainment; migration; attitudes	August 1979	September 1979	October. 1979
Fourth wave: Household compoition, labor-force participation, and income update; child care; higher education; school meals	November 1979	December 1979	.....
Fifth wave: Household composition, labor-force participation, and income update; pension coverage; net worth	February 1980	March 1980	January 1980
Sixth wave: Annual income roundup; job-related benefits; taxes; informal assistance; Infants, Women, and Children benefits; emergency energy assistance	May 1980	June 1980	April 1980

[1] The 1979 Research Panel sample was composed of three samples of equal size. Each third of the total sample was designated as a rotation group to be interviewed on a separate schedule read vertically down each column. The fourth wave was administered to only two-thirds of the sample.

sponses might be less accurate. A portion of the 1979 Panel was interviewed under rules that required self-only interviews except where this was impossible because of illness, prolonged absence, and so forth. Preliminary results indicate that the self-only rules produced more precise reporting at the expense of more interviewer travel and increased losses caused by refusals. Unless analysis currently in progress suggests otherwise, the recommendation for those conducting future surveys will be to choose an intermediate approach that would allow proxy responses and specify key items for telephone callbacks.

Two other, more specialized respondent tests were carried out in connection with the 1979 Panel. The fourth interview wave included a set of items dealing with postsecondary educational enrollment and expenses. Students away at school usually were treated as absent household members and interviewed by proxy. On the assumption that many parents are poorly informed about students' income and expenses, the students themselves were interviewed in most cases. The other test involved leaving a dropoff form, to be returned by mail, to obtain self-employment income. Records of such income are often kept elsewhere than at home, and sometimes by other persons, such as accountants. It was assumed that this procedure would yield more accurate reports of this poorly measured type of income. Procedural problems unfortunately marred this test.

A final area of controlled experimentation involved the frequency of interviews. Though quarterly interviews were usually used, it was acknowledged that longer reference periods would reduce the burden on respondents and lower the cost of collecting income, employment, and household composition data for a calendar year. A 6-month period was tried in the Site Research Test, and for half the 1979 Panel, assets income was collected on a 6-month basis. Result suggest that a reduction in the accuracy of reporting and the burden of longer interviews are not offset by savings in resources and interview time.

Initial Findings of 1979 Panel

Overlapping Receipt of Benefits*

Though they are concerned to a significant degree with the methodological and feasibility tests already described, the results of the 1979 Panel are of interest because they provide a first look at some of the issues that the ISDP was intended to measure. For example, preliminary results are now available on the overlapping receipt of benefits. The first wave of the 1979 Panel provides considerable data on income sources and participation in noncash programs. The file on each adult in the sample contains a recipiency flag for the following sources of cash income: Wage and salary earnings, earnings or draw from own business, earnings or draw from own farm, income from some 15 types of assets, payments from six types of educational assistance from the public and private sectors, and income from any of approximately 30 other sources, including the OASDI program, SSI, AFDC, General Assistance, four types of public and private pensions, veterans' payments, unemployment compensation, workers' compensation, and alimony or child support. Information is also available on participation in such noncash programs as Medicare, Medicaid, food stamps, public housing, and rent subsidies.

Tables 2, 3, and 4 present the initial results on the receipt of multiple benefits by households under the OASDI program, federally administered SSI, public assistance, unemployment compensation, and the food stamp programs. The data are for the 3-month reference period at the beginning of calendar year 1979. The covered population consists of 77.6 million households in the 50 States and the District of Columbia; persons in institutional living arrangements are excluded.

These tables were programmed using preliminary household weights provided by the Bureau of the Cen-

Table 2.—Households receiving benefits from one or more of five major programs in early 1979 [1]

Source of benefits	Total number of beneficiary households (in thousands)	Percentage distribution: Total	Percentage distribution: Benefits under one program	Percentage distribution: Benefits under more than one program [1]	Approximate standard error of percentage
Social Security program	21,917	100.0	83.8	16.2	0.8
Food stamps	5,234	100.0	16.6	83.4	1.9
Unemployment compensation	4,154	100.0	72.3	27.7	2.3
Federally administered SSI	3,615	100.0	23.3	76.7	2.6
Public assistance [2]	3,295	100.0	16.8	83.2	2.4

[1] Each wave of the 1979 Panel has a fixed reference period of 3 months. The overall panel, however, was divided into three equally sized, independent subsamples interviewed at monthly intervals beginning in February 1979. Thus, the calendar reference months for each subsample are overlapping but not the same. For the February 1979 subsample, the reference months are November and December 1978 and January 1979; for the March subsample, December 1978 and January and February 1979; and for the April subsample, January, February, and March 1979.

[2] Public assistance includes AFDC payments, General Assistance, emergency assistance, and other cash welfare payments received from State or local welfare departments, excluding State-administered Supplemental Security Income payments. In early 1979, about 60,000 individuals were receiving State-administered but not federally administered payments. About 200,000 other recipients of State-administered SSI payments were concurrently receiving federally administered payments and are included in that row of the table.

* This section is based substantially on work done by Denton Vaughan, Clarise Lancaster, and Charles Lininger of the Income Survey Development Program Staff.

Table 3.—Patterns of multiple receipt of benefits under five major programs [1] in early 1979 [2]

[Household units]

Source and pattern of benefits [3]	Number of households (in thousands)	Approximate standard error	Percentage distribution
Households receiving one or more of the five types of benefits	30,025	509	100.0
Only one type of benefit	23,642	464	78.7
OASDI only	18,375	414	61.2
Unemployment compensation only	3,005	176	10.0
Food stamps only	868	111	2.9
SSI only	842	109	2.8
Public assistance only	552	88	1.8
Two or more types of benefits	6,383	332	21.3
Only two	4,726	278	15.7
Food stamps and public assistance	1,585	152	5.3
OASDI and SSI	1,128	127	3.8
OASDI and food stamps	527	86	1.8
OASDI and unemployment compensation	459	80	1.5
Unemployment compensation and food stamps	379	96	1.3
SSI and food stamps	361	71	1.2
OASDI and public assistance	199	52	.7
Other combinations	88	30	.3
Three or more types of benefits	1,657	156	5.6
Only three	1,507	148	5.1
OASDI, SSI, and food stamps	773	105	2.6
OASDI, food stamps, and public assistance	210	54	.7
SSI, food stamps, and public assistance	189	51	.6
Unemployment compensation, food stamps, and public assistance or OASDI	193	45	.6
OASDI, SSI, and public assistance	142	44	.5
Four or more types of benefits	150	46	.5
OASDI, SSI, food stamps, and public assistance	90	35	.3
Other combinations	60	25	.2

[1] Cash payments from Old-Age, Survivors, and Disability Insurance (OASDI), federally administered Supplemental Security Income (SSI), public assistance, unemployment compensation, and food stamps. Other benefit programs were excluded; had any of them been included, multiple program participation would have increased.

[2] See footnote 1, table 1.

[3] See footnote 2, table 1.

sus that yield a total household count approximately 3 percent higher than the corresponding count from the March 1979 CPS. The higher count is now generally taken to reflect a bias in the preliminary ISDP weights. An alternative definition of the household weight, based on the weight assigned to the wife of the household reference person in cases where both spouses are present, reduces the total number of ISDP households to near the CPS count. This alternative weighting scheme does not materially affect the estimates in tables 2–4.

The standard errors included in tables 2 and 3 are approximations based on assumptions regarding the numerical relationships between variances for the 1979 ISDP Panel and the 1979 March CPS. The approximations also include adjustments made to reflect the differences in sample design for the two surveys. Direct estimation of the 1979 ISDP Panel variances may lead to substantially different estimates. Results of this work, currently in progress, will be available soon.

Table 2 presents the number of households that received each of the five selected types of benefits as well as the percentage of each that also received benefits from one or more of the other selected programs. It shows that more than four-fifths of the households receiving food stamps or public assistance, and more than three-fourths of those receiving SSI payments, also received at least one of the other types of benefits under consideration. The table also reveals that far smaller proportions of the households with OASDI benefits (16 percent) or unemployment compensation (28 percent) were also in receipt of at least one of the other selected benefits.

Table 3 provides a detailed picture of the patterns of benefit recipiency. About 30.0 million households—or more than 1 out of every 3 in the United States—received benefits from one or more of the five selected programs during the 3-month reference period in 1979. Of the beneficiary households, 23.6 million received only one type of benefit, and 6.4 million received benefits from two or more of the five programs. One measure of the quality of these data is that the ISDP recipient counts are generally superior to those obtained by the March CPS, the best alternative data source available for the study of multiple participation.

Table 4 compares the aggregate estimates of program participation obtained by the 1979 Research Panel with appropriate benchmark data from program sources. It reveals little or no underreporting of OASDI (97.0 percent) and SSI (101.4 percent) and a substantial improvement over the previous "state of the art" reporting of

Table 4.—Comparison of ISDP recipient counts with benchmark data for January 1979

Program	Preliminary benchmark estimate (in thousands)	Preliminary ISDP estimate (in thousands)	ISDP estimate of percent of benchmark
Social Security [1]	28,348	27,489	97.0
Federally administered SSI	3,826	3,878	101.4
Public assistance [2]	[3]4,208	[4]3,371	80.1
Unemployment compensation [5]	3,185	2,931	92.0
Food stamps [6]	5,394	4,656	86.3

[1] Cash recipients aged 18 and older.

[2] AFDC, General Assistance, emergency assistance, and other cash payments received from the State or local welfare departments, excluding State-administered SSI.

[3] Cases.

[4] Adults receiving assistance. For two-parent families nominally only one parent is included in the count.

[5] Includes recipients under the regular State program, as well as ex-servicemen, civilian Federal employees, and persons continuing to receive payments under extended-duration programs.

[6] Units.

unemployment compensation (92.0 percent). The reporting levels for public assistance (80.1 percent) and food stamps (86.3 percent) are comparable, if not higher, than those regularly obtained by the March CPS, despite an apparently depressed count of households headed by women in the ISDP. Since estimates for both public assistance and food stamps are dependent on complete coverage of these households, the counts for both public assistance and food stamps may rise somewhat if sample weights can be revised to account for biases in coverage. This problem is currently under study by analysts in the Department of Health and Human Services and the Bureau of the Census. This process is not likely to lead to an upward revision of the public assistance count by more than 10 percent and should affect the food stamp count to a lesser extent.

Furthermore, the ISDP has made progress in understanding the problems of correctly classifying AFDC recipients and the general assistance population. Additional procedural changes suggested by the ISDP research are expected to lead to improved classification of these programs.

Assets and Food Stamp Eligibility*

The content and timing of the second wave of interviews were adjusted slightly to meet the data requirements of the Food and Nutrition Service (FNS) of the Department of Agriculture. That agency had received a congressional mandate to study the impact of restrictions on the amount of assets held by households eligible for food stamps. Funding and, especially, time limitations made it difficult to accomplish this objective with a free-standing survey, and no established data source was found that could provide the necessary information. Because the 1979 Panel sample overrepresented less-affluent households and because the ISDP questionnaire content had already been designed to collect extensive data on assets and eligibility criteria, it was possible for the 1979 Panel to supply FNS with the required data from a sufficiently large sample to make acceptable national estimates. Preliminary results were presented to Congress in a January 1981 report entitled "Assets of Low Income Households: New Findings on Food Stamp Participants and Nonparticipants." Tables 5, 6, and 7 summarize some of these findings.

Two questions were addressed initially:

(1) What are the assets of food stamp recipients and how do they compare with the assets of nonrecipients?

(2) How many persons are excluded from the food stamp program solely because they have too many assets?

* This section is substantially based on a summary prepared by Jonathan Lane of the Income Survey Development Program Staff.

Table 5.—Assets of food stamp participants,[1] by dollar value, 1979

Dollar value of assets reported	Percentage distribution of households: All assets [2]	Countable assets [3]
Total	100.0	100.0
$0	36.8	48.7
$1-$500	39.0	41.9
$501-$1,000	9.5	5.2
$1,001-$2,000	6.6	1.3
$2,001-$3,000	2.8	1.2
$3,001-$5,000	[4] 2.7	[4] 1.5
$5,001 or more	[4] 2.6	[4] .2

[1] Food stamp participants any time during 3-month reference period.
[2] Excludes home.
[3] Distinction between countable and noncountable assets reflects statutory requirements in effect May–July 1979, when data were collected. Countable assets exclude home equity and specified assets not readily liquidated or those needed for work.
[4] Asset holdings at end of reference period.

An overwhelming majority of food stamp recipients were found to have very few assets. Table 5 shows that if the value of homes is excluded, 37 percent of the households receiving food stamps in 1979 had no assets and 92 percent had total assets of $2,000 or less. When the examination is restricted to "countable" assets (which means, in most cases, excluding the value of homes, personal effects, an automobile, and perhaps some tools of a trade), 49 percent of the food stamp households had no such assets, and 97 percent had countable assets of $2,000 or less.

Thus, the response to the first question is that the vast majority of food stamp recipients have very few assets.

Table 6.—Total dollar value of assets,[1] by food stamp eligibility and recipiency status of household, 1979

Dollar value of assets reported	Percentage distribution of households: Food stamp recipients	Eligible nonrecipients	Ineligible— Solely because of too many assets	Ineligible— Because of too much income
Total	100.0	100.0	100.0	100.0
$0	36.8	30.1	0	3.4
$1-$500	39.0	28.2	0	7.2
$501-$1,000	9.5	13.3	0	5.3
$1,001-$1,500	4.8	7.6	0	4.8
$1,501-$2,000	1.8	6.1	2.6	5.0
$2,001-$3,000	2.8	6.3	8.1	8.1
$3,001-$5,000	2.7	4.0	21.5	14.0
$5,001-$10,000 [2]	2.6	2.9	39.6	25.3
$10,001-$75,000	0	1.1	23.4	25.0
$75,001 or more	0	.5	4.8	1.9

[1] Excludes home.
[2] Specific values in excess of $5,000 were not collected for some types of assets. Nonspecified amounts in excess of $5,000 are assigned to the $5,001-$10,000 bracket.

Table 7.—Type of reported assets, by food stamp eligibility and recipient status of household, 1979

Type of asset	Percent of households reporting asset			
	Food stamp recipients	Eligible nonrecipients	Ineligible— Solely because of too many assets	Ineligible— Because of too much income
Liquid	49.5	53.9	96.9	91.2
Automobile	48.6	58.5	75.5	87.9
Second automobile	11.0	25.1	45.4	52.2
Home	35.5	42.2	68.2	70.2
Life insurance	8.8	14.7	27.3	44.5
Farm	0	1.2	5.3	1.4
Nonfarm business	0	1.4	2.4	4.5
Rental property	2.5	2.8	8.7	6.9
Other	0	.5	1.9	2.3

In response to the second question, estimates from the survey show that about 5.1 million households (containing about 12.3 million persons) were excluded from participation in the food stamp program because of the asset limitation. If these households had not been ineligible because of the asset test and had participated at about the same rate as those who were then eligible (approximately 65 percent), about 8 million more persons would have been in the program in 1979, at an additional cost of about $2.9 billion. Clearly, the asset screen significantly reduces the size of the program.

Perhaps surprisingly, most of the persons excluded from the program because of assets exceeding the stated limit had substantially more, not just a little more, assets than those who participated. Table 6 shows that, among those made ineligible solely by the asset test, nearly 70 percent had assets of more than $5,000, well above the $1,500 eligibility limit for individuals and young families (those headed by persons under age 60).

In table 7 the proportion of food stamp households reporting any assets is compared with that of other households. This table also indicates the breadth of asset coverage in the survey. The eight categories of assets measured were:

Liquid. Checking accounts, savings accounts, stocks, bonds, mutual fund shares, and certificates of deposit.

Automobile. Cars, trucks, and vans.

Home. Equity in owned homes (that is, net of mortgage debt).

Life insurance. Policies with cash surrender or loan values.

Farm. Ownership.

Nonfarm business. Ownership.

Rental property. Equity in housing, apartment, condominium, commercial, industrial, and "other" property.

Other. Interest in farm or nonfarm business (other than own), equity in undeveloped land, royalties, mortgages, personal loans, trusts, estates, and any other assets that bring in money.

Other Uses

In addition to these studies, other uses of the 1979 Panel data are planned, in progress, or have been completed. Midway through the life of the 1979 Panel, for example, Congress mandated an emergency energy assistance program for the coming winter. Questions about recipients and nonrecipients of the new assistance were added for use in the following spring, and the data thus obtained are now being examined for the Office of Management and Budget (OMB). Panel data were also used in a partial updating of the Social Security Administration's 1974 Survey of the Low-Income Aged and Disabled. Other content areas covered by topical modules intended for specific users are foster care (Congressional Budget Office), informal care (OMB), child care (National Institutes of Health), the Women, Infants, and Children program (FNS), and school meals (Congressional Budget Office, General Accounting Office, House Agriculture Committee, and FNS). The Bureau of Labor Statistics has also expressed interest in using ISDP data on income and labor-force activity to examine the issues of labor-force related hardship recommended by the National Commission on Employment and Unemployment Statistics (the Levitan Commission).

As additional data become available to interested users, the 1979 Panel will shed light on many aspects of the present tax and transfer system that are poorly understood. Changes in these programs are likely and are a focus of intense legislative and executive concern. The ISDP has been largely successful in accomplishing its mandate to develop a mechanism for collecting a comprehensive data base that can be used to address these issues.

☆ U.S. GOVERNMENT PRINTING OFFICE: 1982-361-182:117

ATTACHMENT B

The 1984 SIPP Panel Sample

The 1984 Panel Sample

The 1984 SIPP Sample consists of approximately 25,000 designated households selected from a subset of Current Surveys (CS) old and new construction reserve measures (updated through March 1983); the sampling frame is also supplemented by several coverage improvement frames. The SIPP Primary Sampling Unit (PSU) sample is a subsample of the CS A design PSU Sample, or more specifically, the National Crime Survey PSU Sample. Some SIPP PSU's are combinations of CS PSU's; most consist of just one CS PSU. There will be 45 SIPP PSU's which are self-representing; that is, each of these PSU's is treated as a separate stratum and is included in the sample with certainty. The remaining CS A design strata were combined to form 129 nonself-representing (NSR) SIPP strata. One NSR sample SIPP PSU was selected from each of these 129 strata for a total of 174 sample SIPP PSU's.

SIPP strata were formed by combining CS strata having A Sample PSU's with similar proportions of nonwhite persons (1970), urban persons (1970), and families with income below poverty level (1969). A CS stratum was selected from each SIPP stratum with probability proportional to its size. The CS PSU's in the selected CS strata are the NSR sample SIPP PSU's.

Sample living quarters were selected from within each sample PSU at a rate such that all sample living quarters have equal probability of selection.

The SIPP Sample will use address segments with an expected size of around two living quarters. The expected address-segment size will vary by SIPP PSU. Area, special place, permit, and coverage improvement segments will have the usual expected four living quarters per segment. Each segment in the sample will be assigned a rotation code (R=1, 2, 3, or 4), designating the month of initial interview--October, November, December, or January.

ATTACHMENT C

Differences Between 1984 Panel Design

and

Subsequent SIPP Panels

Differences Between 1984 Panel Design and Subsequent SIPP Panels

The 1984 Panel of the SIPP is slightly different from subsequent panels in the numbering and scheduling of interviews. First, the 1984 Panel begins with approximately 25,000 designated households while all other panels begin with approximately 20,000 households. The reduction in sample size occurs because of sample redesign research which indicated that acceptable coefficients of variation for important survey variables can be achieved at the reduced sample size. Second, three-fourths of the households in the 1984 Panel have nine waves of interviewing, all other panels have only eight waves. Third, the 1984 Panel begins in October while all other panels begin in January. As a result, Wave 2 in the 1984 Panel was shortened from four rotation groups to three to allow Wave 5 to begin in January 1985, concurrent with the beginning of the 1985 Panel. This abbreviation was accomplished by moving the fourth rotation group from the end of Wave 2 to the beginning of Wave 3. Consequently, persons in that rotation group will be interviewed one time less throughout the survey.

For most persons using wave files for cross-sectional purposes this change in interviewing schedule will have little effect, assuming the change has no effect on rotation group bias. 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 built from Waves 1 and 2 will have only 75 percent of the sample, or approximately 15,000 households. Similarly, any longitudinal file containing Wave 2 will have 25 percent of the sample with one fewer interview.

ATTACHMENT D

SIPP Interview Schedule (1984-1987)

Figure 1. SIPP Interview Schedule

SIPP INTERVIEW SCHEDULE

MONTH YEAR	1984	1985	1986	1986	ASSIGNED CASES	DATA MONTHS
OCT 83	I				6350	JUNE-SEPT
NOV	I				6350	JUL-OCT
DEC	I				6350	AUG-NOV
JAN 84	I				6350	SEPT-DEC
FEB	U1				5150	OCT-JAN, 1984
MAR	U1				5150	NOV-FEB
APR	U1				5150	DEC-MAR
MAY	U1 P2				5150	JAN-APR
JUN	U2 P2				5150	FEB-MAY
JUL	U2 P2				5150	MAR-JUNE
AUG	U2 P2				5150	APR-JULY
SEP	U2 W1				5150	MAY-AUG
OCT	U3 W1				5150	JUNE-SEPT
NOV	U3 W1				5150	JULY-OCT
DEC	U3 W1				5150	AUG-NOV
JAN 85	U3 VT	I			10600	SEPT-DEC
FEB	U4 VT	I			10600	OCT-JAN, 1985
MAR	U4 VT	I			10600	NOV-FEB
APR	U4 VT	I			10600	DEC-MAR
MAY	U4 R1	U1 TM			9550	JAN-APR
JUN	U5 R1	U1 TM			9550	FEB-MAY
JUL	U5 R1	U1 TM			9550	MAR-JULY
AUG	U5 R1	U1 TM			9550	APR-JULY
SEP	U5 W2	U2 W1			9550	MAY-AUG
OCT	U6 W2	U2 W1			9550	JUNE-SEP
NOV	U6 W2	U2 W1			9550	JULY-OCT
DEC	U6 W2	U2 W1			9550	AUG-NOV
JAN 86	U6 P1	U3 P1	I		15000	SEPT-OCT
FEB	U7 P1	U3 P1	I		15000	OCT-JAN, 1986
MAR	U7 P1	U2 P1	I		15000	NOV-FEB
APR	U7 P1	U3 P1	I		15000	DEC-MAR
MAY	U7 R2	U4 R1	U1 TM		13950	JAN-APR
JUN	U8 R2	U4 R1	U1 TM		13950	FEB-MAY
JUL	U8 R2	U4 R1	U1 TM		13950	MAR-JUNE
AUG	U8 R2	U4 R1	U1 TM		13950	APR-JULY
SEP		U5 W2	U2 W1		8800	MAY-AUG
OCT		U5 W2	U2 W1		8800	JUNE-SEP
NOV		U5 W2	U2 W1		8800	JULY-OCT
DEC		U5 W2	U2 W1		8800	AUG-NOV
JAN 87		U6 P2	U3 P2	I	14250	SEPT-DEC
FEB		U6 P2	U3 P2	I	14250	OCT-JAN, 1987
MAR		U6 P2	U3 P2	I	14250	NOV-FEB
APR		U6 P2	U3 P2	I	14250	DEC-MAR
MAY		U7 R2	U4 R1	U1 TM	13200	JAN-APR
JUN		U7 R2	U4 R1	U1 TM	13200	FEB-MAY
JUL		U7 R2	U4 R1	U1 TM	13200	MAR-JUNE
AUG		U7 R2	U4 R1	U1 TM	13200	APR-JULY
SEP			U5 W2	U2 W1	8800	MAY-AUG
OCT			U5 W2	U2 W1	8800	JUN-SEPT
NOV			U5 W2	U2 W1	8800	JULY-OCT
DEC			U5 W2	U2 W1	8800	AUG-NOV

I - Initial interview

U - Update/number

P1 - Personal history (1) (Marital history, fertility, migration)

P2 - Personal history (2) (Health and disability, work history, education history)

W1 - Assets and wealth

W2 - Assets and wealth update

VT - (Variable topical module- child care, duration of welfare)

TM - Unassigned topical module

R1 - First annual roundup (annual income, taxes, educational enrollment)

R2 - Second annual roundup (annual income, taxes, educational enrollment)

ATTACHMENT E

Links to Administrative Records and the Validation of Social Security Numbers

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. Household 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, which are a function of past earnings histories and retirement decisions of currently ineligible workers rather than the characteristics of current recipients.

In order 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 SIPP control card. 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 SSN's 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 Technical, Conceptual, and Administrative Lessons of the Income Survey Development Program, M. David (ed.), Social Science Research Council, New York). At the conclusion of each month's interviewing during the first interview wave of the SIPP, a special extract file is prepared by the Census Bureau for the Social Security Administration. This file contains a small number of key variables in a format appropriate for machine validation of survey reported SSN's--SSN, name, date of birth, sex, and race. The Social Security Administration then identifies (by machine validation) incorrectly reported SSN's; the SSA also clerically resolves these cases as well as cases for which the survey respondent failed to report their

SSN. This work should be completed in time for the fourth wave interview, at which time a field followup is conducted to obtain missing SSN's (provided they are not "refusals" or "don't have") and to reconcile inconsistencies in SSN or demographic data generated by the computer match or the clerical resolution.

All original sample persons are eligible for SSN validation. This includes all individuals, including children who either report a social security number or who indicate they have a number but cannot supply it.

All persons who enter sample by virtue of their association with original sample people who are eligible for SSN validation are handled differently from the original sample people. New members of the panel (nonsample persons) starting with Wave 2 are held for computer validation until the start of the next panel. Consequently, nonsample persons from Waves 2 through 5 of the 1984 Panel are held and submitted for computer validation with Wave 1 of the 1985 Panel. Likewise, nonsample persons from Waves 6 through 8 of the 1984 Panel and Waves 2 through 4 of the 1985 Panel would be held and submitted for computer validation with Wave 2 of the 1986 Panel.

Finally, to emphasize its importance to interviews, the social security number is identified as a critical data item of the survey requiring completion. These efforts should result in improved accuracy of the survey reported social security numbers, thereby facilitating linkages to administrative record systems.

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 record data. A Census Bureau committee is, however, discussing the topic; recommendations concerning the uses of administrative records in the SIPP are forthcoming.

ATTACHMENT F

Sources of Income Covered in SIPP;

Preliminary SIPP Definitions of Household and Family Income

INCOME SOURCE LIST

INCOME LIST

Code	Type	Code	Type
1	Social Security	28	Child support payments
2	U.S. Government Railroad Retirement pay	29	Alimony payments
3	Federal Supplemental Security Income (SSI)	30	Pension from company or union
4	State Supplemental Security Income (State administered SSI only)	31	Federal Civil Service or other Federal civilian employee pensions
5	State unemployment compensation	32	U.S. Military retirement pay
6	Supplemental Unemployment Benefits	33	National Guard or Reserve Forces retirement
7	Other unemployment compensation (Trade Adjustment Act benefits, strike pay, other)	34	State government pensions
8	Veterans compensation or pensions	35	Local government pensions
9	Black lung payments	36	Income from paid-up life insurance policies or annuities
10	Worker's compensation	37	Estates and trusts
11	State temporary sickness or disability benefits	38	Other payments for retirement, disability or survivor
12	Employer or union temporary sickness policy	40	G.I. Bill/VEAP education benefits
13	Payments from a sickness, accident or disability insurance policy purchased on your own	50	Income assistance from a charitable group
20	Aid to Families with Dependent Children (AFDC, ADC)	51	Money from relatives or friends
21	General assistance or General relief	52	Lump sum payments
22	Indian, Cuban, or Refugee Assistance	53	Income from roomers or boarders
23	Foster child care payments	54	National Guard or Reserve pay
24	Other welfare	55	Incidental or casual earnings
25	WIC (Women, Infants and Children Nutrition Program)	56	Other cash income not included elsewhere
27	Food stamps		

ASSET LIST

Code	Type
100	Regular/passbook savings accounts in a bank, savings and loan or credit union
101	Money market deposit accounts
102	Certificates of Deposit or other savings certificates
103	NOW, Super NOW or other interest-earning checking accounts
104	Money market funds
105	U.S. Government securities
106	Municipal or corporate bonds
107	Other interest-earning assets
110	Stocks or mutual fund shares
120	Rental property
130	Mortgages
140	Royalties
150	Other financial investments

SIPP Preliminary Income Definitions

Income. For each person in the sample 15 years old and over, questions are asked about the amount of money income received each month from a variety of separately identified sources. (The property questions are designed to collect one amount covering all 4 months in the reference period; one-fourth of this amount will be allocated to each month in the period.) The monthly income amounts from each separate source will appear on SIPP data files individually and will also be aggregated into four categories of personal income (defined below): 1) earnings, 2) property, 3) other, and 4) total. Monthly family income from these four categories will be derived by summing the amount within a category for all persons in the same family. Monthly household income will be similarly derived for all persons in the household. The four categories of income are:

1) Total Monthly Earnings. The sum of monthly amounts for 1) pay received (before deductions) from up to two employers including tips, bonuses, overtime pay, commissions, and special Armed Forces pay such as housing allowances; 2) income received from up to two self-employment sources; i.e., one's own businesses, professional practices, and farms. This amount includes salary received and money taken out of the business, practice, or farm for personal or family use. For businesses grossing less than $1,000 per year, only the amount earned after operating expenses is collected; the amount for all 4 months is allocated equally to each month in the reference period. It may possibly include pay from incidental or casual jobs on an irregular basis; e.g., babysitting, lawn mowing. A decision on whether this type of income will be included in the earnings or "other" income category has not been made.

2) Total Monthly Property Income. The sum of monthly amounts for: 1) interest earned on savings accounts, money market deposit accounts, certificates of deposit and other savings certificates, checking accounts (such as NOW and Super NOW), money market funds, United States Government securities, municipal and corporate bonds, mortgages held by the sample person, and other interest-earning assets; 2) dividends earned on stocks and mutual fund shares; 3) rental income cleared after expenses, which may be a negative amount if money was lost on properties; and 4) income received from royalties and/or other financial investments.

3) Total Monthly Other Income. The monthly sum of money income received from all other sources not included as earnings or property income (defined above). It includes income from sources such as Social Security, public and private assistance or welfare, unemployment and disability compensation, alimony, pensions, and annuities. For a complete list of the income sources included in this category, see income codes 1-56 on the Income Source List. (Possible changes in this category still under consideration are the exclusion of food stamp values and lump sum payments and the inclusion of casual and incidental earnings in the "earnings" category.)

4) Total Monthly Income. The sum of earnings, property, and other income as defined above.

ATTACHMENT G

SIPP Wealth Module Data Items

SIPP Wealth Module Data Items

INTEREST EARNING ASSETS

Assets

1. Regular/passbook savings accounts
2. Money market deposit accounts
3. Certificates of deposit or other savings certificates
4. NOW, Super NOW, or other interest earning checking accounts
5. Money market funds
6. U.S. Government securities
7. U.S. Savings Bonds (type E and EE)
8. Municipal or corporate bonds
9. Other interest-bearing assets

Information collected for each asset

A) For asset held jointly with spouse
   Ownership of asset (Y/N)
   Balance as of the last day of the reference period
   *Call back later (Y/N)

B) For assets held in own name
   Ownership of asset (Y/N)
   Balance as of the last day of the reference period
   *Call back later (Y/N)

C) For assets held with other adults
   Ownership of asset (Y/N)
   Share of balance as of the last day of the reference period
   *Call back later (Y/N)

D) For assets held with children
   Ownership of assets (Y/N)
   Balance as of the last day of the reference period
   *Call back later (Y/N)

*Asked only if asset amount is not provided.

DIVIDEND EARNING ASSETS

Asset

1. Stocks and mutual funds

Information collected

A) For assets held jointly with spouse
   Ownership ~~of asset~~ (Y/N)
   Market value as of the last day of reference period
   Dividend income received during reference period
   Any debt against asset (Y/N)
   Amount of debt
   *Call back later (Y/N)

B) For assets held in own name
   Ownership of asset (Y/N)
   Market value as of the last day of reference period
   Dividend income received during reference period
   Any debt against asset (Y/N)
   Amount of debt
   *Call back later (Y/N)

IRA/KEOGH ACCOUNTS

Asset

1. IRA accounts

2. KEOGH accounts

Information collected for each asset type

Ownership of asset (Y/N)
Types of assets held
Number of years contributed to accounts
Balance in accounts as of the last day of the reference period
Contributed to account in last 12 months (Y/N)
*Call back later

* Asked only if asset amount is not provided.

OWN HOME

Asset

1. Own home

2. Mobile home

Information collected for each asset

A) For own homes
   Home owned or being bought (Y/N)
   Current market value of home
   *Year home purchased
   *Total purchase price excluding closing costs
   Mortgage, deed of trust, or other debt on home (Y/N)
   Current amount owed on mortgage or debt
   *Original amount of mortgage or debt
   *Year mortgage obtained
   *Term of mortgage
   Second mortgage or other debt on home (Y/N)
   Current amount owed on mortgage or debts
   *Original amount of loan
   *Year loan obtained
   *Term of loan
   **Call back later

B) For mobile homes
   Mobile home owned or being bought (Y/N)
   Current market value of mobile home
   *Year mobile home built
   *Mobile home manufacturer
   *Length of home
   Mortgage, loan, or other debt on mobile home (Y/N)
   Current amount owed on mortgage or loan
   *Original amount of mortgage or loan
   *Year mortgage or loan taken out
   *Term of loan
   *Monthly mortgage or loan payments (excluding taxes or insurance)
   Mobile home site owned or being bought (Y/N)
   Current market value of site
   Mortgage, deed of trust, or other debt on the site (Y/N)
   Current amount owed on mortgage or loan
   *Original amount of mortgage or loan on site
   *Year mortgage or loan taken out
   *Term of loan on site
   *Monthly mortgage or loan on site (Y/N)
   Second mortgage or loan on site (Y/N)
   Current amount owed on mortgage or loan on site
   **Call back later

* Used for imputation. Asked only if amount is not provided.
** Asked only if market value or loan amount is not provided.

UNSECURED LIABILITIES

Liabilities

1. Store or credit card bills

2. Doctor, dentist, hospital, or nursing home bills not covered by insurance

3. Debt to private individual not in household

4. Loans from banks, credit union, or financial establishment (excluding mortgages, vehicle loans, money owned to brokers, and educational loans)

Information collected for each liability

A) For liabilities owed jointly with spouse
   Existence of debt (Y/N)
   Amount owed as of the last day of reference period
   *Call back later

B) For liabilities owed in own name
   Existence of debt (Y/N)
   Amount owed as of the last day of reference period
   *Call back later

EDUCATIONAL LOANS

Liabilities

1. Education loans

Information collected

Existence of debt (Y/N)
Number of educational loans
All loans held in own name (Y/N)
Amount owed as of the last day of reference period
*Call back later

* Asked only if information or loan amount is not provided.

MOTOR VEHICLES

Liabilities

1. Licensed car or truck (excluding recreational vehicles and motorcycles)

Information collected

A) For asset held by any member of the household
   Ownership of asset (Y/N)
   Number of vehicles owned
   Who owns the vehicles

B) For 3 vehicles owned
   Year, make, and model of vehicles
   Money owed on vehicle (Y/N)
   Amount owed
   *Call back later

RECREATIONAL VEHICLES

Asset

1. Motorcycle, boat, snowmobile, recreational vehicle, or other

Information collected

A) For assets held by any member of household
   Ownership of vehicle (Y/N)
   Number of vehicles
   Who owns the vehicles
   Type of vehicle (motorcycle, boat, snowmobile, recreational vehicle, other)

B) For 2 vehicles
   Market value in present condition
   Money owed on vehicle (Y/N)
   Amount owed
   *Call back later

* Asked only if information or loan amount is not provided.

RENTAL PROPERTY

Asset

1. Rental property or real estate (other than own home).

Information collected

A) For properties owned jointly with spouse
   Own property/real estate jointly (Y/N)
   Number of properties owned
   For 2 properties and a catchall
   Type(s) of property (vacational home, undeveloped land, rental property, commercial property, other)
   Current market value of property
   *Year property purchased
   *Purchase price excluding closing costs
   *Location of property (State; county)
   Any mortgage on property (Y/N)
   Current amount owed on mortgages or debts
   *Year mortgage obtained
   *Original amount of mortgage or debt
   *Term of mortgage
   **Call back later

B) For properties owned in own name
   Own property/real estate (Y/N)
   Number of properties owned
   For 2 properties and a catchall
   Type(s) of property
   Current market value of property
   *Year property purchased
   *Purchase price excluding closing costs
   *Location of property (State; county)
   Any mortgage on property (Y/N)
   Current amount owed on mortgages or debts
   *Year mortgage obtained
   *Original amount of mortgage or debt
   *Term of mortgage
   **Call back later

C) For properties held jointly with others
   Own property/real estate (Y/N)
   Number of properties owned
   For 2 properties and a catchall
   Type(s) of property
   Current market value of property
   Share of property
   *Year property purchased
   *Purchased price excluding closing costs
   *Location of property (State; county)
   Any mortgage on property (Y/N)
   Current amount owed on mortgages
   *Year mortgage obtained
   *Original amount of mortgage or debt
   *Term of mortgage
   **Call back later

* Used for imputation. Asked only if market value or loan amount is not provided.
** Asked only if amount is not provided.

CHECKING ACCOUNT

Asset

1. Cash on hand and checking accounts

Information collected

1. Cash on hand and share of balance in checking accounts as of the last day in reference period
*Call back later

BUSINESS/FARM EQUITY

Asset

1. Business equity

2. Farm equity

Information collected for each asset

Total value of business
Total amount of debt owed on business
Percent of business owned
*Call back later

RETIREMENT PLANS

Asset

1. Pension or retirement plan

Information collected

Employer or union has pension or retirement plan (Y/N)
Covered by plan (Y/N)
Time necessary to be covered by plan
Why not covered by plan
Covered by more than one retirement plan (Y/N)
Make payments toward plan (Y/N)
Employer contribute to plan (Y/N)
Time covered by plan (yr/months)
Vested rights in plan (Y/N)
Time in plan necessary for vested rights
Number of persons employed at this location
Number of persons employed at all locations

* Asked only if asset amount is not provided.

LIFE INSURANCE

Asset

1. Life insurance

Information collected

Covered by life insurance (Y/N)
Type of insurance (term/cash value)
Amount of term insurance coverage (face value)
For 2 life insurance policies with cash value
  Amount of coverage (face value)
  Current amount of cash value
  Ever withdraw savings (Y/N)
  Amount of savings taken out
 *Year purchased
 *Name of insurance company

* Asked only if cash value is not provided.

ATTACHMENT H

Schedule for Administering Fixed and Variable Topical Modules in the 1984 and 1985 SIPP Panels

Table 10.--Households By Monthly Cash Income and Selected Sources of Income and Noncash Benefits
(Monthly Average for Calendar Quarter)

Sources of money income and noncash benefits	Number	Monthly cash income									Median monthly cash income		Mean monthly cash income	
		Under $300	$300 to $599	$600 to $899	$900 to $1,199	$1,200 to $1,599	$1,600 to $1,999	$2,000 to $2,999	$3,000 to $3,999	$4,000 and over	Value	Standard error	Value	Standard error
All households														
RECEIVED SELECTED SOURCES OF MONEY INCOME AND NONCASH BENEFITS														
Earnings														
Property income														
Pension, disability or survivor income														
Social Security or Railroad Retirement														
Private pension														
Federal government														
U.S. Military														
State or local government														
Veteran's														
Other														
Other transfer income														
Private support payments														
AFDC or other cash assistance														
SSI														
Unemployment compensation														
Noncash benefits (means-tested)														
Food stamps														
WIC														
Free or reduced price school meals														
Public or subsidized housing														
Energy assistance														
Medicaid														

Table 9.--Households By Number of Programs in which Members Participate
(Monthly Average for Calendar Quarter)

Means-tested programs recipiency status	Number (Thous.)	Mean monthly income of household				With one or more persons receiving unemployment compensation	
		Cash only		Cash plus near-cash transfers		Total	
		Value	Standard error	Value	Standard error	Number	Percent
All households							
Received one or more means-tested noncash transfers							
Received one							
Food stamps							
WIC							
Free or reduced price school meals							
Public or subsidized housing							
Medicaid							
Received two							
Food stamps and Medicaid							
Medicaid and free or reduced price school meals							
Other combinations							
Received three							
Food stamps, Medicaid, and free or reduced price school meals							
Food stamps, Medicaid, and public or subsidized housing							
Other combinations							
Received four							
Did not receive means-tested noncash transfers							

Repeat same as "All households" for:

Married-couple family households
Family households with female householder, no husband present, own children under 18 years
Other family households
Nonfamily households
Households with householder 65 years or over

Table 4.--Persons By Monthly Earnings By Selected Characteristics
(Monthly Average for Calendar Quarter)

Selected characteristics	Total		Monthly earnings									Mean earnings	
	Number	Percent	Under $300	$300 to $599	$600 to $899	$900 to $1,199	$1,200 to $1,499	$1,500 to $1,799	$1,800 to $2,399	$2,400 to $2,999	$3,000 and over	Value	Standard error
Persons 16 years old and over with earnings		100.0											
RACE AND SPANISH ORIGIN													
White		100.0											
Black		100.0											
Spanish origin		100.0											
AGE													
16 to 19 years		100.0											
20 to 64 years		100.0											
20 to 24 years		100.0											
25 to 34 years		100.0											
35 to 44 years		100.0											
45 to 54 years		100.0											
55 to 64 years		100.0											
65 years and over		100.0											
LABOR FORCE ACTIVITY STATUS													
With some labor force activity													
With a job entire month													
Worked each week													
Full-time worker													
Part-time worker													
Absent one or more weeks without pay													
Spent time on layoff.....													
With a job part of month													
Spent time looking for work or on layoff													
No job during month													
Looked for work or on layoff entire month													
Looked for work or on layoff part of month													
With no labor force activity													
HOUSEHOLD RELATIONSHIP													
Spouse in married-couple family													
Family householder, no spouse present													
Other family member													
Not a family member													
OCCUPATION													
Civilian job or business													
Managerial and professional specialties													
Executive, admin., and managerial													
Professional specialties													
Technical, sales, and admin. support													
Technicians and related support													
Sales													
Admin. support, including clerical													
Service occupations													
Private households													
Service occupations, exc. private household ...													
Precision production, craft, repair													
Operators, fabricators, laborers													
Machine operators, assemblers													
Transportation and material moving													
Handlers, equipment cleaners, helpers													

Repeat same as "Persons 16 years old with earnings" for:

Males 16 years old with earnings
Females 16 years old with earnings

Table 2.--Persons By Mean Household Income and Program Participation Status

(Monthly Average for Calendar Quarter)

Selected characteristics	Number (Thous.)	Mean monthly income of household				Residing in households in which one or more persons received benefits from means-tested programs							
		Cash only		Cash plus near-cash transfers		Total		Cash benefits		Noncash benefits			
										Total		Food stamps	
		Value	Standard error	Value	Standard error	Number	Percent	Number	Percent	Number	Percent	Number	Percent
All persons													
RACE AND SPANISH ORIGIN													
White													
Black													
Spanish origin													
AGE													
Under 16 years													
16 to 24 years													
25 to 34 years													
35 to 44 years													
45 to 54 years													
55 to 64 years													
65 years and over													
HOUSEHOLD RELATIONSHIP													
Spouse in married-couple family													
Family householder, no spouse present													
Other family member													
Not a family member													
WORK DISABILITY STATUS													
Persons 16 to 64 years old													
With a work disability													
No work disability													
REGION													
Northeast													
North Central													
South													
West													
TYPE OF RESIDENCE													
Inside metropolitan areas													
Inside central cities													
Outside central cities													
1,000 or more population													
Inside central cities													
Outside central cities													
Under 1,000 population													
Inside central cities													
Outside central cities													
Outside metropolitan areas													

Repeat same as "All persons" for:

Males
Females

Table 1.--continued
p. 2

TYPE OF RESIDENCE

Inside metropolitan areas .....
Inside central cities .....
Outside central cities .....
1,000 or more population .....
Inside central cities .....
Outside central cities .....
Under 1,000 population .....
Inside central cities .....
Outside central cities .....
Outside metropolitan areas .....

LABOR FORCE ACTIVITY STATUS

With some labor force activity .....
With a job entire month .....
Worked each week .....
Full-time worker .....
Part-time worker .....
Absent one or more weeks without pay .....
Spent time on layoff.....
With a job part of month .....
Spent time looking for work or on layoff .....
No job during month .....
Looked for work or on layoff entire month .....
Looked for work or on layoff part of month ....
With no labor force activity .....

Repeat same as "All persons" for:

Males
Females

Table 1.--Persons By Monthly Household Cash Income By Selected Characteristics

(Monthly Average for Calendar Quarter)

Selected characteristics	Number	Monthly household cash income									Median monthly household cash income		Mean monthly household cash income	
		Under $300	$300 to $599	$600 to $899	$900 to $1,199	$1,200 to $1,599	$1,600 to $1,999	$2,800 to $2,999	$3,400 to $3,999	$4,000 and over	Value	Standard error	Value	Standard error
All persons														
RACE AND SPANISH ORIGIN														
White														
Black														
Spanish origin														
AGE														
Under 16 years														
16 to 24 years														
25 to 34 years														
35 to 44 years														
45 to 54 years														
55 to 64 years														
65 years and over														
HOUSEHOLD RELATIONSHIP														
Spouse in married-couple family														
Family householder, no spouse present														
Other family member														
Not a family member														
WORK DISABILITY STATUS														
Persons 16 to 64 years old														
With a work disability														
No work disability														
REGION														
Northeast														
North Central														
South														
West														

Note: Universe excludes persons in farm households and group quarters (applies to all tables).

ATTACHMENT O

Examples of Types of Tables Proposed for SIPP Quarterly Reports

Tentative List of SIPP Quarterly Report Tables

Table 1	Persons by Monthly Household Cash Income by Selected Characteristics (Monthly average for calendar quarter)
Table 2	Persons by Mean Household Income and Program Participation Status (Monthly average for calendar quarter)
Table 3	Persons by Labor Force Activity, Status, Mean Household Income, and Program Participation Status (Monthly average for calendar quarter)
Table 4	Persons by Monthly Earnings by Selected Characteristics (Monthly average for calendar quarter)
Table 5	Households by Labor Force Activity, Status, Mean Income, and Program Participation Status (Monthly average for calendar quarter)
Table 6	Household by Monthly Cash Income by Selected Characteristics (Monthly average for calendar quarter)
Table 7	Household by Mean Income and Program Participation Status by Selected Characteristics of Householder (Monthly average for calendar quarter)
Table 8	Households by Receipt of Selected Noncash Benefits and Sources of Money Income (Monthly average for calendar quarter)
Table 9	Households by Number of Programs in which Members Participate (Monthly average for calendar quarter)
Table 10	Households by Monthly Cash Income and Selected Sources of Income and Noncash Benefits (Monthly average for calendar quarter)

ATTACHMENT N

Tentative List of SIPP Quarterly Report Tables

Overview of Weighting for SIPP Quarterly Reports*

This document outlines the procedures that will be used to weight the data for SIPP quarterly reports. The focus is on the first of these reports (third quarter 1983) with some indication of likely differences for subsequent reports. The final weight for each case will be the product of four factors.

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 will simply be the inverse probability of selection. With few exceptions, this factor will be uniform across cases. For subsequent waves, this factor will decrease 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 interviews, information on census region, residence status, race of reference person, tenure (own or rent), and household size will be 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 waves of interview, the household noninterview adjustment factors must account for household refusals and temporarily absent households as well as households whose members all move to an unknown address or to a place more than 100 miles from the nearest SIPP PSU. 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 -- will be used to calculate the household noninterview adjustment factor.

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

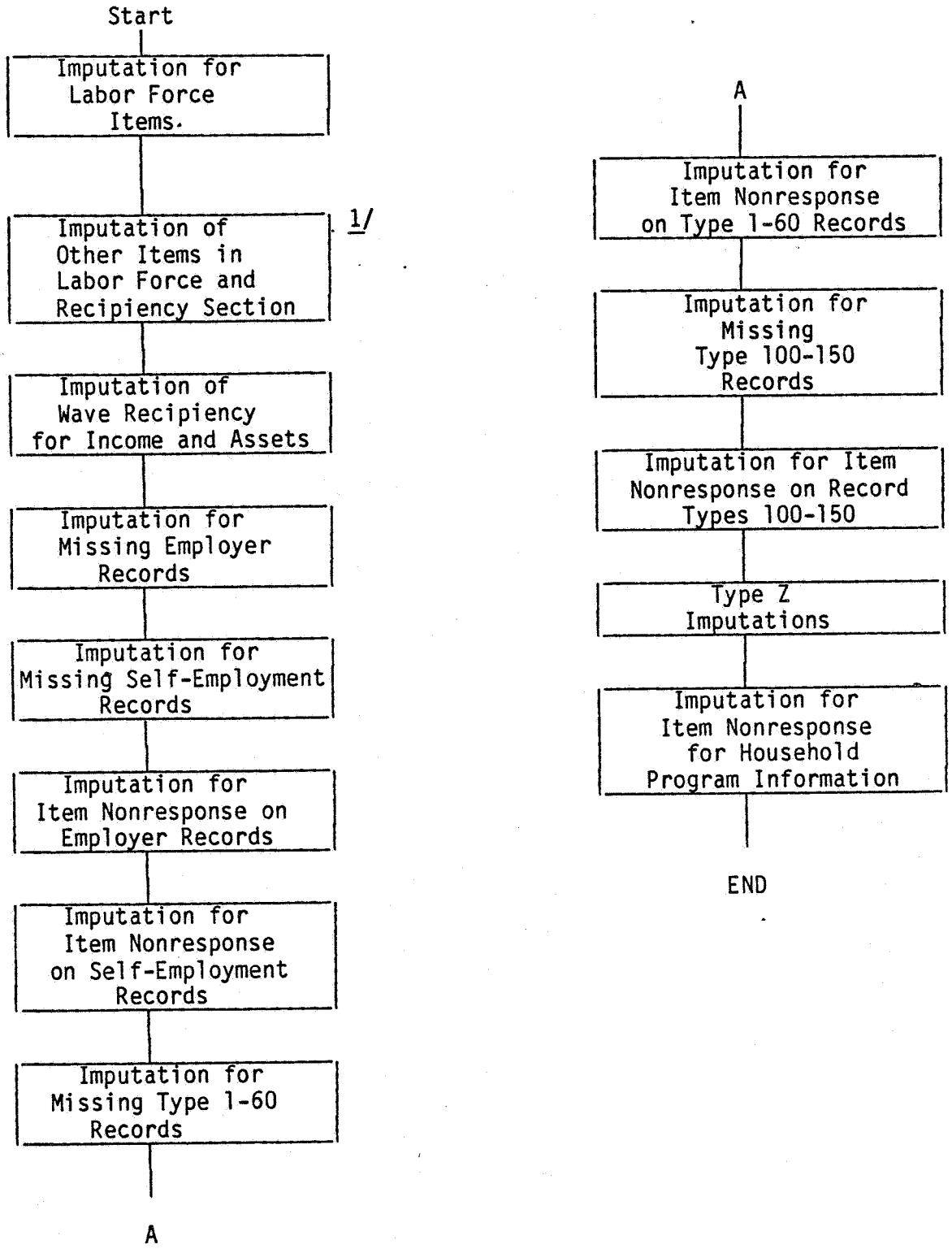
The fourth factor accomplishes several goals simultaneously. It reduces the mean square error of estimates of characteristics of persons correlated with age, race, or sex. Independent current demographic estimates of population by age, race, and sex will be used for this purpose. In addition, monthly estimates of type of household (marital and family status of householders by race and sex) from the Current Population Survey (CPS) will be used to reduce variances on estimates related to these controls. Under this procedure, the monthly estimates will be 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, we will use an iterative raking procedure.

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

ATTACHMENT M

Specifications for SIPP Cross-Sectional Weighting

Preliminary Outline of Cross-Sectional Imputation System for Item Nonresponse



1/ Excludes income and asset recipiency for 4-month period.

Preliminary Cross-Sectional Imputation System for Item Nonresponse

A flowchart of the cross-sectional imputation system for item nonresponse is provided in this attachment. It shows thirteen major steps in the imputation system and indicates the sequence in which the data items will be imputed.

The imputation for item nonresponses will be made using the traditional sequential hot-deck procedure. The data file will be sorted to provide some control for geographical location and processed sequentially. The cells of the hot-deck matrices will then be loaded with data from respondents (one value per cell). The value existing in a specified cell will be 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 will be a two-step operation. First, initial or "cold-deck" values for each cell will be supplied. Second, "live" values from donors should be 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 will also include the development of detailed diary information on the performance of the hot decks and the placement of imputation flags on the microdata file for each item imputed.

ATTACHMENT L

Preliminary Outline of the Imputation System

for

SIPP Cross-Sectional Data Files

Consistency Edit. Performed to ensure 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.

SIPP Edit Procedures

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 data transmissions as they are received in a central location from the regional offices. It is similar to ENTREX editing but is more extensive. The 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 ascending source code order.
6. Checks legality of source codes.
7. Reformats the data into a hierarchically structured file.

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.

Person and Family Edit. Performed to create 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 allocated at this point.

ATTACHMENT K

Brief Description of SIPP Edit Procedures

	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

person number uniquely identifies each person in the SIPP survey, and can be used to link persons across waves. The PSU, segment, serial, and 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. The operational phase makes no attempt to impose longitudinal household definitions based on changing relationships, nor to identify and number longitudinal households.

To illustrate the SIPP identification numbering system, consider the following example. 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. 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.

PSU, segment, and serial number. The PSU, segment, and serial numbers are not changed in later waves, regardless of movers.

The two-digit address ID code is added to provide a means for identifying more than one unique household associated with the same PSU, segment, and serial number. While this situation does not occur in Wave 1, the two-digit address ID code was nevertheless used in Wave 1 to provide consistency for interviewers. The first digit of the address ID code indicates the wave that an address is first assigned for interview. The second digit sequentially numbers multiple households that have the same PSU, segment, and serial number. The Wave 1 address ID code for all sample addresses is 11. As SIPP sample persons move to new addresses, the office staff assigns new address ID codes to each new address. Previous address ID codes 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 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 entry address ID and person number. It is assigned by the interviewer during the listing or updating of the household roster. The interviewer lists the name of each person in the household, and then transcribes the current two-digit address ID code to each person's record. This two-digit number is the entry address ID. 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 the roster in Wave 2; and so forth. The first digit indicates the wave the person enters the survey. This five-digit number consisting of entry address ID and person number is not changed or updated in subsequent interviewers, regardless of movers.

Thus, the fourteen-digit combination of PSU, segment, serial, entry address ID, and

The SIPP Identification Numbering System for Field Control Operations*

The SIPP identification scheme used for field control operations is designed to uniquely identify individuals in each wave, provide a means of linking the same individuals over time, and grouping individuals into unique households each wave. No attempt is made in field operations to define or number each "different" household viewed in a longitudinal way over time. However, all households that form in later waves due to splits can be linked to the original Wave 1 household.

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

The various components of the operational identification scheme are:

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

Census Bureau-Washington staff assign the PSU and segment 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.

Office staff in the twelve regional offices assign the two-digit serial number and the two-digit address ID code. The two-digit serial number is assigned to each SIPP living quarters within a segment. In Wave 1 the combined PSU, segment, and serial number uniquely identifies one sample address. Thus, a SIPP household can be uniquely identified in Wave 1 with these three components:

* Anne Jean, Statistical Methods Division, is the author of this attachment.

ATTACHMENT J

SIPP Identification Numbering System

for Field Control Operations

2.	Limitations	Do not spend more than 5 hours nor travel more than 100 miles while attempting to locate the new address. If you need more time or need to travel further, contact your supervisor.
3.	Record New Addresses	Record the new address(es) in item 34a of the original control card. If you need more space to record new addresses, use the "Notes" space on page 1 of the control card.
4.	Report Address to Office	Report all new addresses to the office immediately. The office will let you know if the case needs to be transferred.
	a. Interview new address	If you are instructed to interview at the new address, prepare a new control card according to the instructions in Chapter C3 of the interviewer manual. Locate the new address, and obtain interviews for <u>all</u> household members who are 15 years old or older.
	b. Reassignment of case	If the new address is assigned to a different interviewer, the office will ask you to provide information from the original control card. Then the office will transfer the case to another interviewer.
	c. Transfer of case	If the office transfers a case from another interviewer to you, record the information provided by the office on a new control card, locate the unit and obtain an interview.
5.	Type D noninterviews	If a mover's new address is more than 100 miles from a SIPP PSU or if a mover's new address cannot be determined, you will report a Type D noninterview.

called "two hundred plus" persons. You may find that additional persons have moved in with sample persons when you update the household roster. Likewise, when you follow a sample person to a new address, you may find that the sample person moved in with persons who are not in the SIPP survey. In both instances, the new persons are added to the household roster.

a. When to follow — If an additional person moves to a new address, he/she is followed ONLY if sample persons have also moved to the same new address. In other words, as you follow the sample person to the new address, you automatically follow the additional persons.

b. When NOT to follow — Do not follow or obtain new addresses for an additional person(s) who moves and is no longer living with any sample persons.

3. Replacement Households — A replacement household is one that has moved into a sample address that was previously occupied by a sample household. The replacement household entirely replaces the original sample household; that is, none of the household members (sample persons and/or additional persons) previously listed on the control card still remain at the address. Do not obtain any control card information for a replacement household. Do not complete any questionnaires for members of a replacement household.

B. General Procedures — Obtain new addresses for all sample persons who move, record the new address in item 34a on the original control card, and report the address to the office.

1. Obtain Complete Addresses — Obtain complete addresses for all sample persons who move. Include as much information as possible, such as apartment designations (if applicable), ZIP codes, directions such as north, south, etc., county name, and State. Verify the spelling of all street names. Complete and accurate addresses are essential and may require some investigation.

a. Inquire — Inquire to find out new addresses. You may ask:

1) Persons who are now living at the original address.

2) Other sources such as mail carriers, rental agents, real estate company, etc.

3) Personnel at the post office.

When inquiring, identify yourself as an interviewer for the Census Bureau and show your identification if it is a personal inquiry. If asked why you need the address, explain that it is necessary to locate the person so you can interview him/her for a survey that is being conducted by the Census Bureau. Do not mention the name of the survey.

b. Telephone directories — A current telephone directory or directory assistance may be able to provide a new telephone number if a person has moved to a nearby address.

INTERVIEWERS INSTRUCTIONS FOR FOLLOWING MOVERS*

A. Movers' Rules

Interviews are conducted every 4 months with the same households (persons) who were interviewed (or should have been interviewed) at the time of the first visit in Wave 1; that is, October, November, December 1983 or January 1984. Households (sample persons) who move are followed and interviewed at their new addresses, within certain limits presented below. The movers' rules apply to all waves except Wave 1.

1. Sample Persons

Those persons who are household members during Wave 1 are sample persons. They are assigned person numbers in the 100 series, such as 101, 102, etc. Sample persons are also referred to as "one hundred series persons."

a. When to follow

If a sample person moves to a new address during the period in which the SIPP is being conducted, efforts are made to locate and interview the sample person and his/her household at the new address. If you discover a mover, attempt to learn the new address, and report it to your office as soon as possible. Your office will tell you if you, personally, will follow up on the interview assignment, or whether the assignment will be given to a different interviewer who lives closer to the new address.

b. When NOT to follow

There are four situations in which a sample person is not followed to a new address. These occur if the sample person is now:

(1) Institutionalized - The institution must be listed in Table A, Specified Institutions, in Part B, Chapter 5 of the 1984 SIPP Panel Interviewer Manual, and the sample person must be an inmate of the institution; that is, he/she occupies a nonstaff unit.

(2) A member of the Armed Forces living in barracks.

(3) Living outside the United States.

(4) Living beyond the limit for following movers. The limit for the 1984 SIPP is 100 miles from a SIPP PSU within the United States. The office compares the new address with the boundaries of nearby SIPP PSU's and determines whether the address is within or beyond this limit. Therefore, you must report each new address to the office so this determination can be made.

2. Additional Persons

Additional persons are those who become household members with sample persons AFTER the initial interview. Since they are assigned person numbers in the two hundreds for Wave 2, the three hundreds for Wave 3, etc., they are occasionally

* Excerpted from Chapter C4 of the 1984 SIPP Panel Interviewer Manual. Detailed instructions for recording movers' data, for example, date entered and left, reason for move, etc, are given in other chapters of the 1984 SIPP Panle Interview Manual.

ATTACHMENT I

Interviewers Instructions for Following Movers

SURVEY OF INCOME AND PROGRAM PARTICIPATION
TOPICAL MODULE SCHEDULE

INTERVIEW DATES	1984 PANEL			198 PANEL		
	Wave	Fixed Topical Module	Variable Topical Module	Wave	Fixed Topical Module	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	Health Care Utilization and Financing Social Services in Health Social Services in Training			
Sept 84- Dec 84	4	Assets Liabilities	Pension and Retirement Issues Housing Conditions and Cost Energy Usage			
Jan 85- Apr 85	5		Child Care Arrangements and Financing Duration of Welfare Support for Non-Household Members Social Services in Child Care and Other Reasons for Not Working-Reservation Wage Work Related Expenses	1	None	None
May 85- Aug 85	6	Annual Income Taxes Employee Benefits Education Enrollment	Educational Financing Training Questions (ETA)	2	Education Enrollment	Educational Financing Training Questions (ETA)
Sept. 85 Dec. 85	7	Assets Liabilities	Housing Conditions and Cost Neighborhood Conditions Energy Usage	3	Assets Liabilities	Housing Conditions and Cost Neighborhood Conditions Energy Usage
Jan. 86- Apr 86	8	Marital History Fertility Migration	Support for Non Household Members Reasons for Not Working-Reservation Wage	4	Marital History Fertility Migration	Support for Non-Household Members Reasons for Not Working - Reservation Wage
May 86- Aug 86	9	Annual Income Taxes Employee Benefits Educational Enrollment	Educational Financing Health and Disability Work Related Expenses Retirement Plans and Expectations Training Questions (ETA)	5	Annual Income Taxes Employee Benefits Educational Enrollment	Educational Financing Health and Disability Work Related Expenses Training Questions (ETA)
Sept 86- Dec 86				6	Assets Liabilities	Open
Jan 87- Apr 87				7	Health and Disability Work History Education History	Open
May 87- Aug 87				8	Annual Income Taxes Employee Benefits Education Enrollment	Educational Financing Training Questions (ETA)

WORKING PAPER NO. 01

"An Overview of the Survey of Income and Program Participation"
D. NELSON, D.B. MCMILLEN and D. KASSPRZYK (Census Bureau)

(Includes papers 8401-8407)

8401 "An Overview of the Survey of Income and Program Participation" by Dawn Nelson, Dave McMillen, Daniel Kaspryz

8402 "Toward a Longitudinal Definition of Households"
by David Byron McMillen, Roger Herriot

8403 "Papers presented in the SIPP Session I 8/13-16/84:

"An Analysis of Turnover in the Food Stamp Program"
by Irene Smith Lubitz, Timothy J. Carr

"The Wealth and Income of Aged Households"
by Daniel B. Radner (Social Security Administration)

"The Measurement of Household Wealth in the SIPP"
by Enrique J. Lamas, John M. McNeil

8404 (Not Available)

8405 "The Survey of Income and Program Participation"
by Roger A. Herriot & Daniel Kasprzyk

8406 Papers Presented in the SIPP Session III at the Annual Meeting of the ASA in Philadelphia 8/13-16/84

"Obtaining Cross-Sectional Estimates from a Longitudinal Survey: Experiences of the Income Survey Development Program" by Hertz Huang

"Weighting of Persons for SIPP Longitudinal Tabulations" by David R. Judkins, David L. Hubble, James A. Dorsch, David B. McMillen and Lawrence R. Ernst

"Longitudinal Family and Household Estimation in SIPP" by Lawrence R. Ernst, David L. Hubble, and David R. Judkins

8407

"Month-to-Month Income Recipiency Changes in the ISDP" by J.C. Moore and D. Kasprzyk, Bureau of the Census

"Findingss from the Student Follow-Up Investigation of the 1979 ISDP" by A. M. Roman and D. V. O'Brien, Bureau of the Census

"The ISDP 1979 Research Panel as a Methodological Survey: Implications for Substantive Analysis" by R. A. Kulka, Research Triangle Institute

"Some Data Collection Issues for Panel Surveys with Application to SIPP" by A. Jean and E. K. McArthur, Bureau of the Census

"Managing the Data from the 1979 ISDP" by P. Doyle and C. Citro, Mathematica Policy Research, Inc.

THE SURVEY OF INCOME AND
PROGRAM PARTICIPATION

AN OVERVIEW OF THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

No. 01

D. Nelson, D.B. McMillen and
D. Kasprzyk
Bureau of the Census

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.

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.

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 Ycas 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

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.

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

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.

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	
1	2	FEB	OCT NOV DEC	JAN				
	3	MAR	NOV DEC	JAN FEB				
	4	APR	DEC	JAN FEB MAR				
	1	MAY		JAN FEB MAR	APR			
2	2	JUN		FEB MAR	APR MAY			
	3	JUL		MAR	APR MAY JUN			
	4	AUG			APR MAY JUN	JUL		
3	1	SEP			MAY JUN	JUL AUG		
	2	OCT			JUN	JUL AUG SEP		
	3	NOV				JUL AUG SEP	OCT	
	4	DEC				AUG SEP	OCT NOV	
4	1	JAN				SEP	OCT NOV DEC	
	2	FEB					OCT NOV DEC	JAN
	3	MAR					NOV DEC	JAN FEB
	4	APR					DEC	JAN FEB MAR
5								
6				Continues in the pattern established above.				
7								
8								

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.)

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.

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.

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.

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 <u>and</u> 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.)

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

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

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.

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

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.

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

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 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

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.

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 concept.

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 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.

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.

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.

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

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.

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 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.

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 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.

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.

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

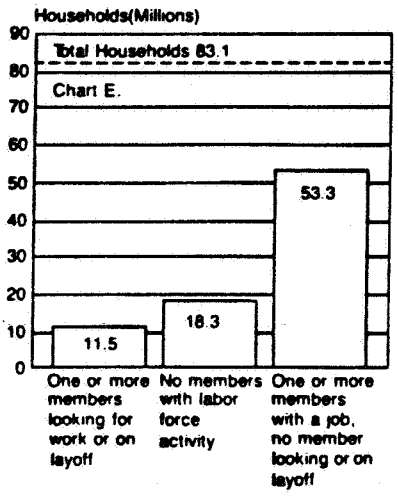
Third Quarter 1985 Report and Wave 7 (1984) Public-Use File.....6/86

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

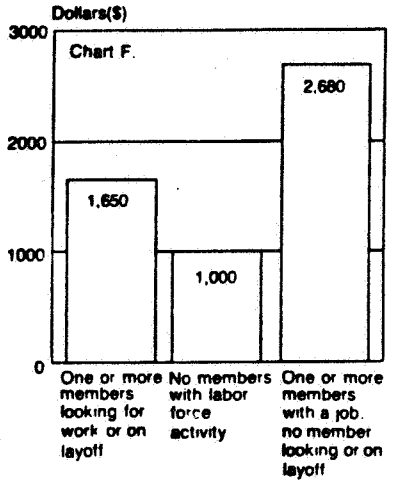
Households, by Labor Force Activity

(Third Quarter 1983)



Mean Monthly Cash Income of Nonfarm Households by Labor Force Activity

(Third Quarter 1983)

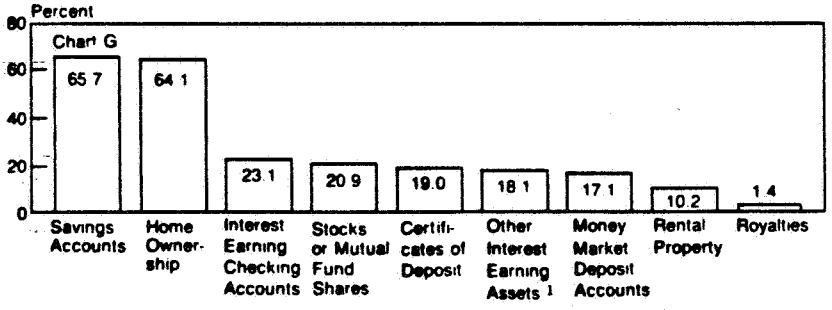


Asset Ownership

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)

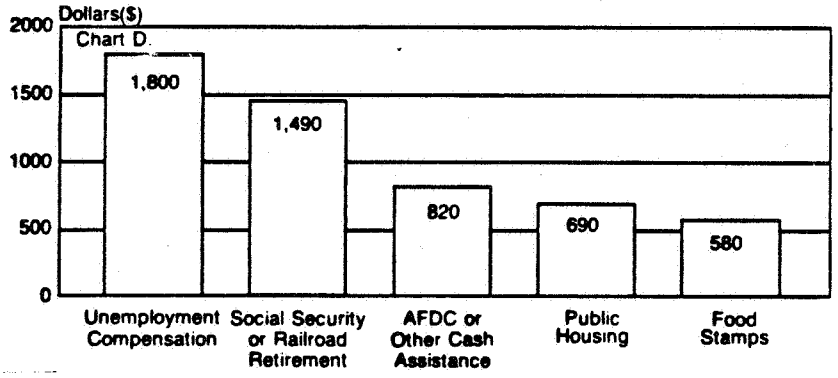


[1]Includes money market funds U.S. Government securities, municipal or corporate bonds, mortgages, and U.S. Savings Bonds.

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.

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)

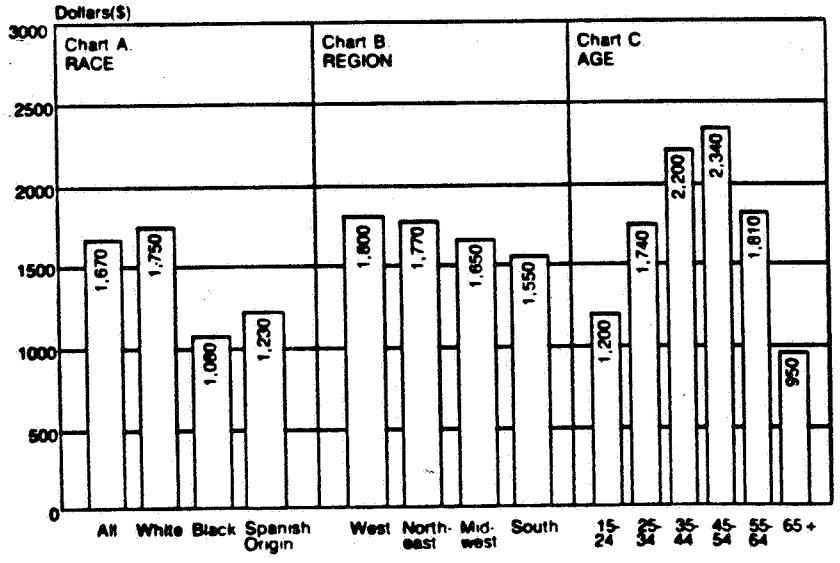
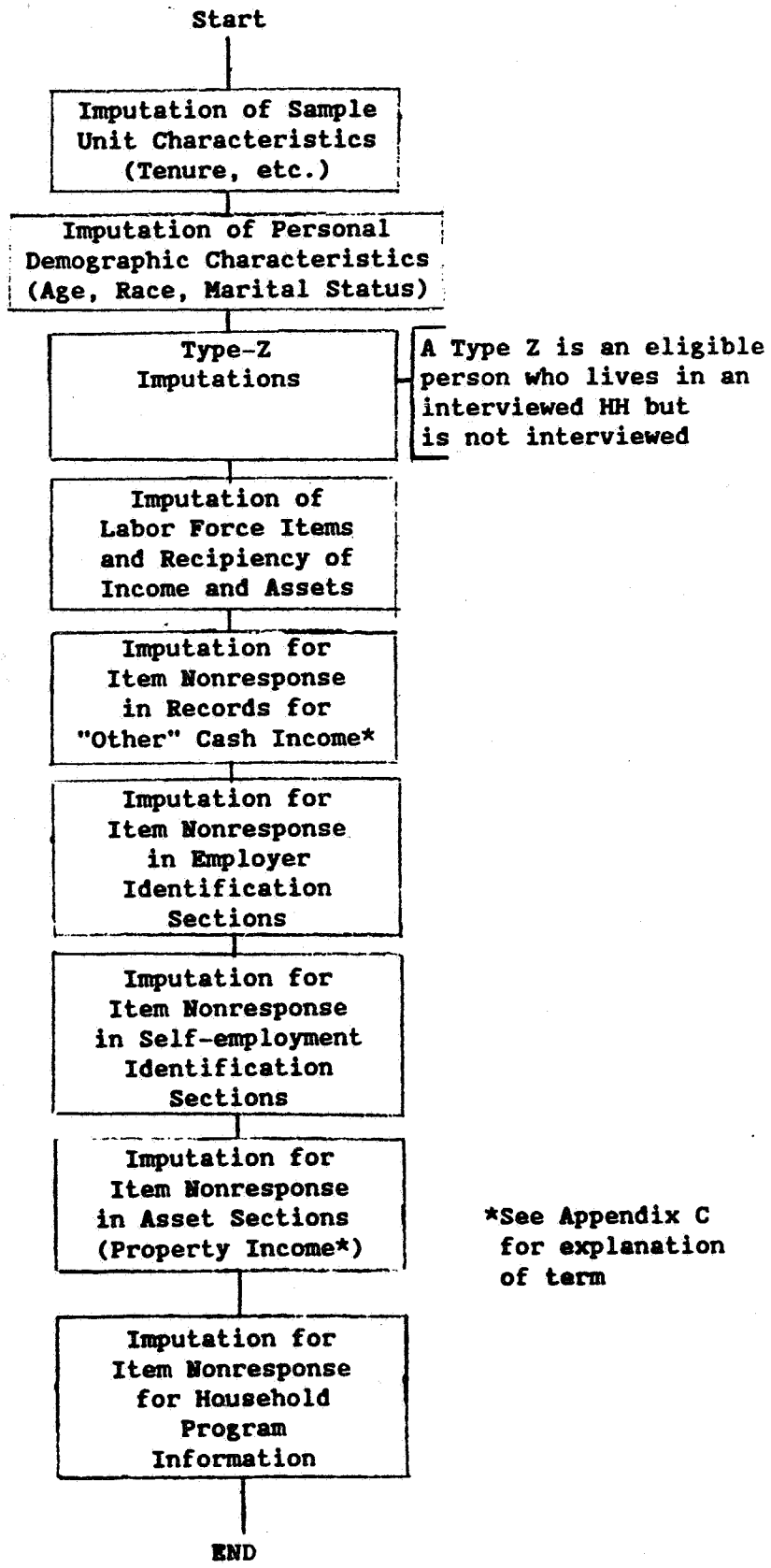


Figure H-1. Cross-Sectional Imputation System Flowchart



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).

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.

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.

Herriot, R.A. and D. Kasprzyk (1984). The Survey of Income and Program Participation, SIPP Working Paper Series No. 8405. 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.

Survey of Income and Program Participation

Working Paper Series

No. 8402

Toward a Longitudinal Definition of Households

by
David Byron McMillen,
and Roger Herriot

Acknowledgments

This paper was prepared by **David Byron McMillen** and **Roger Herriot**, Population Division. This paper has benefited from the comments of a number of our colleagues at the Census Bureau. Special thanks go to **Daniel Kasprzyk** and **Paul Siegel** who have read and commented on innumerable drafts.

Suggested Citation

McMillen, David Byron, and Herriot, Roger. "Toward a Longitudinal Definition of Households." SIPP Working Paper Series No. 8402. U.S. Bureau of the Census, Washington, D.C. 1984.

TOWARD A LONGITUDINAL DEFINITION OF HOUSEHOLDS

Introduction and Background

Data collection and analysis in the social sciences generally focus on cross-sectional surveys such as the Current Population Survey (CPS). Consequently, most of our concepts and data analysis tools are structured around point estimates of some phenomenon or characteristic. To the extent that we try to develop longitudinal concepts and measures of social phenomena, that is to say viewing events across time rather than at one point in time, we conflict with these cross-sectional structures. It is the goal of this paper to confront that cross-sectional/longitudinal conflict and attempt some reconciliation. More specifically, this paper attempts to develop longitudinal definitions of households and families which are useful for observing these units across time and for constructing aggregate characteristics across that time period, while not creating serious conflict with our cross-sectional constructs of household type and composition. We begin this exercise by examining cross-sectional household concepts from the CPS and recounting the deficiencies of that perspective. Next we will examine several types of longitudinal definitions, identify the type that is usually cited as most useful, and describe three definitions within that framework. In the third section of this paper we will evaluate the definition options available in terms of utility as well as what is possible given the data at hand to implement such a definition. Next we will illustrate how this definition might be used in calculating aggregate household characteristics and in tabulations of the number of households, household types, and household characteristics.

Point Estimates and Longitudinal Measures

The household definitions used in the CPS serve as adequate measures for the intended cross-sectional purposes. Indeed, few people argue that these definitions create a problem when estimating the number of households by type at the time of the survey. However, when those definitions are used in conjunction with other variables, problems begin to develop.

Discussions on measuring annual household income from the CPS center around the retrospective nature of the measure. Household members as of March are asked to recall their income for the previous calendar year, and the income of all members are aggregated to create a household income. The problem centers on the varying lengths of household membership and the unvarying time period used in the aggregation. Even though a person is a member of the household for part of the year and, thus, contributes income for only part of the year, that person's entire annual income is included in the household income. Similarly, persons who are not members of the household at the time of the survey are not included in calculating the annual household income, even though they may have contributed income for most of the year. This type of criticism is often used to question the adequacy of the CPS income measures; however, it is better viewed as an example of the problems created by combining a point estimate of household composition with a longitudinal (annual) measure of income. Inevitably, the compromises necessary to combine such cross-sectional and longitudinal constructs produce a less than ideal measure.

Similar criticism of the CPS household data can be made. If we examine consecutive March measures of the distribution of households by type we observe little change. The CPS measure of households masks most of the interesting change in the distribution of households. For example, in recent years the number of married couple households has changed at a rate of less than 1 percent a year, or about 200,000 households. Concurrent with that indistinguishable change are over 3 million marriages and divorces, not to mention changes in household type as a result of the death of one member. The small net change creates the appearance of stability, while masking considerable activity. Again, the problem is not so much the inadequacy of the data, but rather the difficulty of measuring longitudinal events with point estimates.

When criticism is leveled against a particular measure, the problem often is not the measure but rather the incongruity between the measuring instrument and the time frame being considered. The examples used above are annual measures, but the same problem exists regardless of the length of time. Most social measurement is discrete while time is continuous. The goal of course is to get to the point where the difference between the two is trivial and can be easily ignored.

In summary, much of the criticism of CPS measures can be attributed to this discrepancy between the time reference of the social measurement and the cross-sectional survey instrument. One solution to the problem is to decrease that difference by repeatedly measuring the phenomenon in question during the year. Those observations can then be aggregated to produce measures which cover a number of time intervals. It is from this perspective that the design for the Survey of Income and Program Participation (SIPP) has developed.

The design of this survey is to interview the household every 4 months over a 2 1/2-year period, and to collect in those interviews monthly data on household composition, income, labor force participation, and a number of other characteristics. Those monthly data can then be aggregated to larger temporal units such as quarterly or annual measures. However, with the idea of aggregating monthly units comes the problem of defining which units should be aggregated across time and which should not. That is to say, which households are the same over the period, which exist at the beginning of the period but not at the end, and which exist at the end but not at the beginning. Without such a definition, aggregating above the person level is impossible.

Defining households across time is an issue that has been debated for several years without resolve; however, it is necessary that the Census Bureau decide

which of the many proposed methods will be used for the publication series from the Survey of Income and Program Participation (SIPP). This paper moves one step nearer to that decision by summarizing the proposals on how longitudinal households should be defined and recommending a system to be used. In addition, this paper will begin to identify what conceptual and processing problems remain unsolved given the definition chosen.

Several proposals have been offered for defining longitudinal households. Griffith (1978) outlines six measures, one of which is the traditional Current Population Survey (CPS) definition. Griffith also proposes that the Census Bureau use several definitions in tabulating households from SIPP. Others who had proposed variant measures include two from Davey (1980), Crosby (1979), Lane (1978), and Smith-Ycas (1981). Ycas (1981) in summarizing past work identifies four keys for labeling definitional methods: static; dynamic; static-dynamic hybrids; and attribute methods. In the following section we will discuss several types of longitudinal definitions for households.

Types of Longitudinal Household Definitions

Static definitions of households fix the household composition and characteristics at a given point in time and calculate other attributes from that point. These definitions are the standard cross-sectional perspective on households common to the CPS and other similar surveys. Using a point estimate of household composition, other attributes are calculated assuming that the composition chosen existed for the full period. Thus, some estimate of annual income for each member is aggregated to produce an annual household income, regardless of whether members were there for the full period or joined the day before the interview. This type of household definition is the logical outgrowth of cross-sectional surveys where interviews are conducted at one point in time

and aggregates of past events are a function of respondent recall. This type of definition coincides with the instantaneous conception of a household which we use from day to day.

Static definitions are both useful and familiar for cross-sectional surveys; however, they serve little purpose in longitudinal surveys other than to provide familiarity. Static definitions, for a number of reasons, ignore the dynamic activity common to households--households are formed by marriages and dissolved by divorces, children leave home and set up their own household, or move in with relatives, and so on. It is difficult to justify the expense and complexities necessary to measure these dynamics if we then suggest to ignore them in defining households. It is useful to portray static definitions here, however, for they represent one end of the definition spectrum.

Dynamic definitions of households occupy the other end of the spectrum. These definitions recognize change as inherent in observing households across time, and attempt to incorporate that change into the definition. Thus, household characteristics and attributes become variables to be measured as households change, are created, and dissolve during the period of observation. In other words, these definitions attempt to minimize the extent to which dynamic concepts are forced into static categories. Needless to say, dynamic definitions are better suited to a longitudinal survey such as SIPP; however, such definitions are difficult to devise and to carry out.

One of the first difficulties encountered with dynamic definitions is that they produce measures which are not readily familiar to many of those who use census data. The most common illustration of this point uses household size. Static definitions produce measures of household size such as 2 or 3, which are intuitively meaningful. That is, they fit with our instantaneous image of house-

holds because they represent the household size at one point in time--the survey date. Dynamic definitions produce measures of household size which look more like averages across a number of households--2.4 or 3.2 members in the household. These measures are summary statistics of the household experience, summarizing across time. In other words, dynamic definitions force us out of that instantaneous view of households and into thinking about them as something which change across time; our statistics produce a summary of that change. Ycas (1981) suggests several ways of handling the problem of household size-rounding, using modal size, etc.--however, it may be best to reeducate the reader to think of annual household characteristics as the aggregate of a number of discrete experiences. There are other more troublesome problems to be dealt with in developing dynamic household definitions. I will deal here only with definitional problems, acknowledging that there are also measurement problems to be considered later in this paper.

Unlike many demographic variables, there are several aspects of dynamic households for which there is no definition or consensus as to what constitutes a change in type. For example, if a husband and wife divorce, there are several ways we can account for this on our household ledger. We could count this as the dissolution of the husband/wife household and the formation of two new households. This results in a net increase of one household in existence at that point in time and an increase of two households when counting the number of households existing during the period. Alternatively, we could allow one household to be the continuation of the husband/wife household. Again we have a net increase of one household in existence, but because of the continuing household, we increase only by one the number of households existing during the period. To generalize, a household may experience a number of changes across time and we can converse easily about the discrete events. However, we do not have a clear

concept of when those changes result in the formation of a new household and the dissolution of an old household. One extreme is to say that any change to the household composition results in a new household. At the other extreme are those who say that this is an issue without resolution and suggest that we abandon the measurement of household characteristics except as they pertain to individuals. In other words, before we can implement a dynamic definition of households, we must first develop a set of continuity rules or accounting principles which identify cases of household dissolution, household formation, and cases where two households at two points in time are identified as the same household.

Most longitudinal household definitions that have been proposed fall somewhere between the static and dynamic extremes. Each acknowledges the difficulty of developing continuity rules, and proposes some static-dynamic blend to finess those problems. A number of cross-sectional/dynamic hybrid definitions have been proposed. One set of these definitions is quasi-dynamic and acknowledges that a set of continuity rules has yet to be developed. Another set is basically a static system designed to avoid the continuity dilemma. Neither of these alternatives is particularly attractive. In the latter case, most of the alternatives create as many problems as they solve. In the former case, if we are going to develop a set of continuity rules, then there is little need for a hybrid definition.

Attribute-type definitions are drawn from the work done on the Panel Study of Income Dynamics (PSID). The goal in these definitions is somewhat different than in the previous discussion. Rather than attempting a longitudinal definition of households, this system calculates a series of cross-sectional households at some smaller time interval, and then ascribes the characteristics of

the household to each individual. Measures for some longer time interval are then calculated by aggregrating the series of point estimates across each individual to represent that person's household experience during the period. This system will yield the number of persons who lived in "households" with a monthly income of $1000 to $1500 during the year; however, it is more difficult to derive the number of households with an annual household income of $12,000 to $18,000. In fact, without an additional set of assumptions, this system does not produce an accounting of households across time. In order to develop household statistics within the attribute system it is necessary to assume, for example, that the householder at the end of the period will represent the household experience. Then the household attributes ascribed to that person are aggregated to produce household characteristics. Those aggregrated attributes represent the householder's experience during the period, but not necessarily the experience of the other persons in the household at that time. As can be seen, an assumption such as this contains many of the weaknesses of using a static or cross-sectional definition of households, with few obvious advantages at the household level.

In summary, there are four types of household definitions which have been proposed for use with longitudinal surveys: cross-sectional; dynamic; cross-sectional/dynamic hybrids; and the attribute system. The cross-sectional approach is clearly inappropriate since it ignores the dynamic nature of the data. The attribute system incorporates the dynamic aspects of the data but dodges the issue of developing continuity rules for households. Consequently, this system, by ignoring the social structure of households, produces many of the same problems raised in criticism of the CPS measure of annual household income. It is clear that a dynamic definition is the most desirable alternative, but agreement on just how that definition ought to be formed is elusive.

In the following section, I will discuss dynamic definitions of households and present three sets of continuity rules. The first, proposed by Norton (1982), defines change, rather than continuity, as movement between major types of households. The second is a reciprocal majority rule system developed by Dicker and Casady (1982). The third, developed by Siegel (1982), sets out a set of demographic principles to which continuity rules should conform and then develops a continuity rule within that framework.

A Dynamic Definition of Households

A dynamic definition of households is much easier to describe than to execute. Dynamic household definitions do little more than acknowledge change in household composition or type and determine that the change must be incorporated. In other words, dynamic definitions acknowledge a set of accounts and to some extent set up the framework for those accounts, but usually do not explicate the principles by which the ledger is filled.

For SIPP, it is suggested that we tabulate the changes that occur in household composition and type during the period covered by a panel. That is to say, we must acknowledge that households change, are formed, dissolve, and sometimes stay the same. In the simplest form then, our accounts will record the formation and dissolution of households from our original sample as well as changes in size and type. These tabulations of change will use as often as possible the standard descriptors, such as family and nonfamily households, and the categories associated with those types. Our dynamic definition begins with the static CPS definition defined at the beginning of the panel and then traces the changes that occur to those households across the duration of the panel. Norton (1982) suggests that we define a household as changed when its membership changes in such a manner as to classify it as a different type of house-

hold. He proposes that we acknowledge change between the following types of households: 1) married-couple household; 2) male family-household; 3) female family-household; 4) male nonfamily-household; and 5) female nonfamily-household. Thus any change which results in a household falling into a different category results in the dissolution of one household and the formation of another. To illustrate this system consider a husband-wife household which experiences a divorce. Norton's scheme would consider the husband/wife household dissolved and two new households formed. The two new households would be family households if there were children present and nonfamily households otherwise.

A second longitudinal definition has been proposed by Dicker and Casady (1982) for use with the National Medical Care Utilization and Expenditure Survey (NMCUES). In a slight departure from the definitions discussed here, Dicker and Casady focus on families rather than households; however, that does not pose any serious problems. Like others who have approached this problem, Dicker and Casady begin with the realization that there is not consensus on when families begin or cease to exist; rather, such transitions are in part a function of the problem being investigated.

The NMCUES model for defining longitudinal families requires that antecedents and descendent families or, in their terms, predecessor and successor families be defined reciprocally. That is to say, any rules defining relationships across time must be applied to both families simultaneously. Dicker and Casady next demonstrate that when applying these rules you wind up with links between a number of households. That is to say, any family is likely to have more than one predecessor and more than one successor family. Thus, the problem lies in defining which of the possible pairs will be defined as the longitudinal

family. As with most longitudinal definitions, the system eventually reduces the decision to what will be defined as the same and what will be defined as change.

Dicker and Casady chose to define sameness by a majority rule. The successor family which receives the majority of members from the predecessor family is identified as the "principal predecessor." These two families then form the linked or longitudinal family unit. Finally, in cases where families split evenly, it should be noted that the NMCUES model does not define a longitudinal unit, but rather dissolves the predecessor family and considers all successor families as newly formed.

Five rules of relationships focus Siegel's (1982) development of household demography. The first two state that a household can have only one descendent and one antecedent that is identified as the same household. That is to say, when a household splits, only one of the subsequent households can be identified as "the same" household. The third and fourth rules identify households which are not the same as some preceding or succeeding household. Households which are not the same as any antecedent household are newly formed; a household not the same as any descendent household is dissolved. The final rule, one of transitivity, states that if A is the same as B and B is the same as C, then A must be the same as C. All that remains to complete this set of accounting principles is a definition of sameness. The rule is offered that two households separate in time and having the same householder are the same household.

Continuity based on following the householder has been criticized because of the somewhat arbitrary way in which the householder is defined, and because it creates what some consider unreasonable change within a continuing household. The most frequently cited example of such change is following the male after a

divorce when the children remain with the female. An alternative to Siegel's householder rule which is consistent with his demography of households is to follow the principal person. The principal person is the female in the married-couple household and the householder in all other households: this is the concept used in developing household weights in CPS. By following the principal person, we alleviate the problem cited above. Of course the problem now occurs when the children stay with the male following a divorce, a much less frequent event.

FIGURE 1

	Norton	Siegel	Dicker
Time 1	ABcd*	ABcd*	ABcd*
Time 2	A* Bcd*	A Bcd*	A* Bcd
Time 3	Ac* Bd	Ac Bd	Ac Bd
Time 4	Acd B*	Acd B	Acd B

*New household

Let us consider briefly the strengths and weaknesses of these three systems focusing on two issues: 1) the number of households created across time, and 2) the extent to which the definition promotes or discourages longitudinal analysis. Norton's system comes the closest to maximizing change and, as a result, creates more households than the others. Consider the divorce example from above, but two children remain with the female. Both Siegel and Dicker and Casady would produce a total of two households resulting from the divorce. Norton's system produces three households: 1) the original married-couple household; 2) a male nonfamily-household; and 3) a female family-household. Let us continue following these people and assume that the children leave the female one at a time and join the male (see figure 1). In Norton's scheme, the first move by a child would produce the dissolution of the male nonfamily-household and the creation of a male family-household. Our longitudinal count

of households now stands at four. Neither Dicker and Casady nor Siegel would produce new households as a result of the children moving. When the second child moves, the female family-household is dissolved and a female nonfamily-household is created. The male family-household remains unchanged. Over these four observations, Norton's system produces five households; both Siegel and Dicker and Casady produce only two by allowing the continuation of a household across these changes. Let us then look at those continuing households. The continuous household for Siegel's householder rule starts as a four-member married-couple household, dwindles to one member--the male, and increases to two with the addition of one child and then to three with the addition of the second child. On the other hand, Dicker and Casady's continuous household begins as the four-person married-couple household and is transformed by the divorce to a three-person female-headed household, then to a two-person and, subsequently, a one-person nonfamily-household. It should be noted that these two continuous households follow opposite courses after the divorce. The continuous household under the principal-person rule would be identical to Dicker and Casady's continuous household.

We should stop at this point and examine what we can learn from this cumbersome yet realistic example of household change. First, if we are interested in counting households (the number or percent of households in poverty during the year, for example), then a continuity system such as Norton's, which allows for continuity in only the most trivial cases, creates a much larger number of households. Suppose the female half of our mythical household was in poverty after the divorce. By Norton's count, during that year we would have 20 percent of the households in poverty. Dicker and Casady and Siegel would show 50 percent of their households in poverty. A second observation to be made

here is that, regardless of what sort of continuity rule we adopt, we will observe households which contain a wide variety of change. The question we must ask is whether we accept that households undergo such change and remain intact.

As noted above, each of the three systems has its constituency and its detractors. Siegel's system is criticized because of the disjunctures that can occur following a divorce. For example, the continuous household will follow a male householder who divorces his wife even though the wife and children remain in the housing unit as a group. Similarly, Norton's scheme is criticized because of the lack of attention paid to continuity. Dicker and Casady are criticized by the mechanical nature of the majority rule. Why, it is asked, should one person make all the difference in whether a family is designated new or continuous?

None of the definitions of longitudinal households offered in the literature has proved viable. However, in the process of discussing this issue with several demographers and economists, it became clear that any definition which labels a transition from a family household to a nonfamily household as continuous causes problems. Drawing on that experience, we determined that we should develop a longitudinal definition of families separate from that for nonfamily households.

We begin with the CPS definition of a family as two or more persons, one of whom is the householder, related by birth, marriage, or adoption, and residing together. To make this cross-sectional definition dynamic, we must add the time dimension or develop a continuity rule. Thus, a longitudinal family is defined as two or more related persons, at least one of whom is the householder

or spouse of the householder, who had the same household experience over two or more consecutive months. We further stipulate that no more than one core family unit with children can continue from a previous-month family. Three levels of criteria are offered to distinguish cases where both parents and children split into two or more households. The first-level criterion, for continuity, is that the family with the most child-months is identified as continuous. The second level, to distinguish between families with the same number of child-months, is the family with the most family-months. In the third level, if two potential continuing units tie on both of the above criteria, then the continuing unit will be assigned randomly. Two elements have been added to the CPS definition: 1) the time dimension, and 2) the inclusion of spouse as part of the continuity criteria.

Let us examine this definition more carefully. Consider again our four-step example of divorce and then the movement of two children one at a time from one parent to another. Following the separation, the Bcd family would be the continuing family because it contains two or more members of the initial family, one of whom is the householder or spouse. The A household would be new because of the transition from family to nonfamily status. Following the movement of the first child, c, the Ac family is considered newly formed because of the transition from nonfamily to family status. Finally, the movement of d from the Bd family to the Ac family results in a new nonfamily household, B. Using the notation from figure 1, we have:

1.	ABcd*	3.	Ac*	Bd
2.	A* Bcd	4.	Acd	B*

Next, we must confront defining continuity for nonfamily households. A nonfamily household is two or more unrelated individuals residing together or an

individual living alone. For these cases, we have adopted a 50-percent rule. As long as the householder and 50 percent or more of the household is the same at two points in time, the household is considered a continuous household. The distinction between this and the majority rule is that, rather than creating new households for even splits, this rule provides for continuity.

Given the above definitions for longitudinal family and nonfamily household, we have now defined the full set of longitudinal households in SIPP. However, other possible longitudinal units or groups exist in relation to federally funded support programs. For example, food-stamp units and AFDC units are defined independently of the household and, in fact, households may contain more than one of these units. Longitudinal units for these programs will be defined on the basis of the person in whose name the program application is filed. For example, in a husband-wife, two-child family, the male is defined as the food-stamp recipient. If he leaves, that food-stamp unit is dissolved; a new one is formed if the female reapplies and is found eligible.

Perspectives on Household Characteristics

In this section, I will address the uses of this longitudinal definition of households and argue that we need to tabulate household data from SIPP using at least two types of longitudinal definitions. The need for two types of definitions is a function of the kind of household information needed. Before I develop this argument, it is useful to record here an analogy developed by Paul Siegel to illustrate the function of a longitudinal definition.

Consider, if you will, a chemistry research lab with a number of rooms in which radioactive elements are used. This lab also has a number of employees who work with these radioactive materials, and, like any good bureaucracy, the workers find themselves constantly on the move. Each week, when they come to work, their

workroom assignment has changed and they go off to work with yet another radioactive element in yet another room. To complete our analogy, consider the rooms in this lab to be our households and the workers the members of those households. Fortunately, the rooms neither move nor change shape, so sameness isn't a problem here, but there are other aspects which are helpful. As good administrators, we are worried both about the safety of our researchers and staying within the guidelines for the maximum allowable accumulation of radioactivity in a lab. And so we set out to measure radiation levels when it dawns on us that we must take two measurements. First, we must determine the radiation level in each room; since that varies, as does the amount of time each worker spends in each room, that is not enough. To measure the exposure of each worker, we also must measure the time spent by each worker in each room. Our analogy has a fairytale ending; in this case, we were able to rely on administrative records to collect the room-duration data, and our study came in on time and under budget.

A number of useful parallels can be spun from this analogy, but I will confine myself here to a discussion of how we will look at these data. First, our households/rooms are our longitudinal unit, and part of what we are interested in is characteristics of that unit. How many people have been members, or what is the membership of that unit over the year? We also want to characterize the room by the amount of radioactivity (income) collected there and the changes that have occurred within that unit during the year. However, these tables do not tell us much about the well-being of our workers. Thus, we also need to examine individual experiences. What types of rooms did they work in during the year? How can we characterize the accumulated experiences in those rooms?

Going back to our types of longitudinal definitions, I am arguing that we need to use both a dynamic longitudinal and an attribute-type household definition, because we are interested in both the experience of households and of individuals in households. There are some characteristics like annual household income which we need to examine both as a characteristic of the household and of the individual. This is only to say that there are multiple meanings associated with the concept of income. In CPS, where we have only one way of obtaining income data, we attach all of those meanings to that single measure. SIPP allows us to decompose that measure and look at the components more carefully.

To summarize, I have argued that to fully appreciate the household dynamics we observe in SIPP and to portray that activity over a year, we should provide two types of tabulations. The first tabulates household characteristics using a longitudinal definition and examines how changes in some characteristics result in changes in others. The second type of tabulation examines how household characteristics affect individuals across time.

It is useful at this point to make one other comment about dynamic household tables. Regardless of the definition used, for any given period of time, we will have one set of households which existed for the full period and one set which did not exist for the full time period. For a number of reasons, we should tabulate these groups separately. The first group has a full set of information and present no problem. The second group, those households newly formed or dissolved during the year, are quite different. First, they do not have data for the full year. Second, we know that they have experienced some change--how drastic depends on the definition chosen--and, thus, we are likely to want to examine different items for these cases. Finally, we cannot calculate annual characteristics for part-year households without imposing a set of assumptions.

Calculating Household Income

Regardless of the definition chosen, many households will be continuous yet experience some demographic change. It then must be specified whose income to include for what periods of time. Thus, even though we define most households as the same, those households are likely to experience some compositional change. It is that compositional change that complicates the calculation of an annual household income. For households that remain the same but experience some compositional change, we propose that income be aggregrated from the monthly responses across all members of the household for the duration of the membership for the full year. Thus, if two persons were members of a household for the full year and a third person with income joined for 3 months, a total of 27 person-months of income would be summed to produced an annual household income. We propose a similarly straight-forward method for calculating the average household size. The household size is the sum of person-months divided by the number of months in the period (12 for years; 3 for quarters, etc.). In the case above, the annual household size would be 2.25. This procedure for calculating household income has a number of advantages. First, it accumulates income only for those persons in the household and thus avoids the inflation/deflation problems inherent in a static definition. It also maintains the short recall advantages built into SIPP. Other advantages include the conceptual clarity of the definition, as well as some processing advantages which will be discussed later.

While the calculation of annual household incomes for households which remain the same is rather simple, some problems remain with defining such income for households which exist for only part of the observation period. Using the procedure above, we can calculate aggregate annual household income for households that are born or die during the period; however, it is not clear that

these incomes should be treated as the same information as that calculated for households which existed for the full year. To illustrate some of the problems, let us take a household with one member there for 8 months with an income of $500 per month, and a second member present for 4 months with an income of $600 per month. The aggregate income for that household is $6400; however, the annual income is somewhat between that minimum and $13,200 which assumes that both household members earn at that rate for a full 12 months. An additional problem is raised in the calculation of household size. For this example, the household would have a annual size of 1 (12 person-months/12) unless we propose some other rule for defining the denominator. Using the greatest household duration for all members would yield a household size of 1.5 (12 person-months/8); however, it would not be comparable to an annual household size. The income problem is somewhat less problematic for households which dissolve than for households that are formed. Dissolved households will earn no more income, and, thus, we have complete information for that household. Newly formed households are likely to have their income truncated because they did not exist at both time points being measured. One solution for both the income and size problems is to tabulate separately those households that remain the same and those households newly formed or dissolved during the observation period. Some alternate measure from annual income, such as average monthly income, or an annualized income could then be computed for those households. Similarly, household size could be calculated using the number of months the householder was present in the denominator and person-months in the numerator.

Another option for tabulating part-year households is to weight them by their duration. Thus, a household existing for 6 months is counted as one-half of a household; a household existing for 9 months is counted as three-fourths of a household; and so forth. However, it is still an issue as to which cell of the

annual income these households should be tabulated. To tabulate a one-half-year household with 6 months of income totalling to $6000 as a household with a $12,000 annual income, a constant flow of income would have to be assumed. Similarly, to tabulate that household as one-half a household with an annual income of $6000 is also misleading. A more accurate representation of this situation is to tabulate it as one-half a household with an average monthly income of $1000. This incorporates both the time dimension of the household and the recorded income information without distorting the tabulation of income. If, on the other hand, what you want is a tabulation of annual income, the solution should be to tabulate only those households that are defined to exist as a continuing unit over the full period and to clearly note that this represents a limited or restricted universe of households. If this option is pursued, some research should be initiated to investigate whether any biases are introduced by censoring the data in this manner.

It should be noted here that, while annualization is mentioned as an option, it is included only because it is an option many have proposed. I contend that annualization is the worst possible solution. Our goal is to measure the income received by a household over a given period of time. Much of the effort of SIPP is designed to give as accurately as possible a picture of the changing membership of those households across time. To compute an annualized income for a household that is newly formed during the period, it would have to be assumed that the household existed unchanged across that period; we know that is not true. As noted above, we must live with the dilemma that for some households it is not possible to compute annual aggregrates. It is possible, however, to compute subannual measures for nearly all households (at least for households that exist for one month or more). For example, the monthly rate of income can be computed for all households and is a reasonable base of

comparison. However, to compare an income flow measure with an annual income stock creates more problems than it solves.

Longitudinal Household Weighting

Having developed a longitudinal household definition and having begun to lay out the ways in which we want to aggregate and tabulate household data, it is now time to confront the issue of developing longitudinal weights for the households in SIPP. The target population of households used for these weights is all households that exist during the 2 1/2 years covered by the SIPP panel. Our longitudinal definition allows us to identify all households in existence at the beginning of the survey and continuing throughout, as well as all households that are newly formed from those households or dissolved. All of these households can be linked to the population from which the initial sample was drawn. There is, however, a set of households which are formed from persons outside the sample. Anyone residing overseas, in an institution, or in an armed forces barracks at the time of the first interview who subsequently leaves that special population will not be captured by this survey until they come in contact with a person or household in the sample. Any households formed by persons from one of those special populations will not be captured by this survey. Without an adjustment for the omission of these newly formed households, we will consistently underestimate the number of households after Wave 1.

Drawing on the experience of the National Center for Health Statistics (NCHS), it is suggested that weights be developed only for households containing a person from the Wave 1 sample at the beginning of the survey. Weighting of other households would require an extensive set of assumptions. In other words,

weights will be developed for all Wave 1 households and all new households formed from those households.

Processing Longitudinal Households

By calculating annual household income, we assume that we have a fully linked file for the calendar period in question. However, it is possible to proceed with either a hierarchical or relational data base. To calculate annual household income from a hierarchical data base would be easiest if all persons who at any time were members of a household were grouped together. This could result in duplicate records for some persons, or it would mean that antecedent and descendent households would need to be contiguous. In addition, each person record would have to carry some set of identifiers indicating what months that person belonged to which household. In other words, before computing the annual income measures, it would be necessary first to link all records so that we had information for the first five waves. Second, households defined as the same, newly formed, or dissolved must be identified and the links between newly formed and antecedent households must be identified. Third, the data set should be sorted so that each same household is followed by its descendent households. Thus, when processing for annual income, we will know that, when we begin to process a household which is unrelated to the previous household, we have processed all cases for that previous household and all of its descendents. Finally, separate tallies can then be made for full year annual income and for annualized income for newly formed and dissolved households.

REFERENCES

J.E. Griffith. Impact of a Longitudinal Survey on the Data Base. (ISDP Working Paper #1, 1978.)

E. Davey. Memorandum for the record. 12/24/80.

B. Crosby. Memorandum to Lane, Ycas, and Mahoney. 7/79.

J. P. Lane. Four Alternative Ways of Defining "Annual" Analysis Units. (ISDP Working Paper #6, 1978.)

Smith-Ycas. Developed by Ycas based on reports of proposal by J. Smith. 1981.

M. A. Ycas. Income Analysis Units for Longitudinal Files. 1981.

A. J. Norton. Notes prepared for 11/18/82 meeting of SIPP Household Task Force.

P. Siegel. Notes prepared for the 10/81 meeting of the Census Advisory Committee on Population Statistics.

M. Dicker and R. J. Casady. A Reciprocal Rule Model for A Longitudinal Family Unit. Paper presented at the 1982 meetings of the American Statistical Association.

Survey of Income and Program Participation

Working Paper Series

Papers Presented in the
Survey of Income and Program Participation
Session I
at the Annual Meeting of the
American Statistical Association
in Philadelphia, PA
August 13-16, 1984

(No. 8403)

U.S. Department of Commerce
Malcolm Baldrige, Secretary
Clarence J. Brown, Deputy Secretary
Sidney Jones, Under Secretary for
Economic Affairs

BUREAU OF THE CENSUS
John G. Keane,
Director

Working Paper Series No. 8403

Session: Survey of Income and Program Participation I

Session Chair: Paul Planchon
Mathematica Policy Research, Inc.

Papers:

"An Analysis of Turnover in the Food Stamp Program."
Written by T. Carr and I. Lubitz, Mathematica Policy Research, Inc.

(Analyses turnover in participation and the eligibility to participate in the food stamp program.)

"The Measurement of Household Wealth in SIPP."
Written by E. J. Lamas and J. M. McNeil, Bureau of the Census.

(Discusses the measurement of wealth in the ISDP, imputation, and early SIPP results.)

"The Wealth and Income of Aged Households."
Written by D. B. Radner, Social Security Administration.

(Discusses the economic resources of aged households using data on both income and wealth.)

Other papers presented in this session were unavailable for publication.

Preface

This report is comprised of three papers featured in the "Survey of Income and Program Participation" Session I in the Social Statistics Section of the Annual Meeting of the American Statistical Association.

These sessions covered a range of topics, both methodological and substantive, about longitudinal surveys and the Survey of Income and Program Participation (SIPP).

SIPP is a new Census Bureau survey collecting data that will help measure income distribution and poverty throughout the country more accurately. These data will be used to study Federal and state aid programs (such as food stamps, welfare, Medicaid, and subsidized housing), to estimate future program costs and coverage, and to assess the effects of proposed changes in program eligibility rules or benefit levels.

Households in the survey will be interviewed at 4-month intervals over a period of 2 1/2 years. The reference period will be the 4 months preceding the interview. In all, about 20,000 households will be interviewed, approximately 5,000 each month. Field operations will be handled through our 12 regional offices.

Recurring questions will deal with employment, types of income, and noncash benefits. Periodic questions will be added dealing with school enrollment, marital history, migration, disability, and other topics. Special supplemental questions will also be added to the SIPP questionnaire.

These papers discuss SIPP and its predecessor, the Income Survey Development Program (ISDP), an experimental program designed to test procedures used in conducting SIPP.

AN ANALYSIS OF TURNOVER IN THE FOOD STAMP PROGRAM

Irene Smith Lubitz
Timothy J. Carr

July 1984

The findings reported in the paper are based on research carried out by Tim Carr, Pat Doyle and Irene Smith Lubitz as a part of MPR's analysis of participation in the Food Stamp Program, pursuant to contract No. 53-3198-0-101 with the Food and Nutrition Service, USDA. The authors are indebted to Steven Carlson, Harold Beebout, Jim Ohls and others for helpful comments on the research on which this paper is based. The conclusions presented here are solely those of the authors and do not necessarily reflect the opinions of the Food and Nutrition Service or Mathematica Policy Research.

INTRODUCTION

Researchers and policymakers have long had an interest in analyzing the factors associated with the decision of eligible individuals and households to participate (or not to participate) in income maintenance programs. Accordingly, a large and growing body of research on this topic has appeared in recent years. There has also been considerable interest in patterns of program participation over time (i.e., entry into, and exit from, these programs). However, there have been relatively few studies of the longitudinal patterns of participation (which we hereafter refer to as "turnover"), because of stringent data requirements. Such studies that have appeared have usually been based on peculiar samples (e.g., participation in negative income tax experiments).

The Survey of Income and Program Participation (SIPP) will provide an unprecedented opportunity for policy analysts to study turnover in transfer programs. In this paper we report the findings of an analysis of turnover in the Food Stamp Program based on data from the 1979 Income Survey Development Program Research Panel (hereafter referred to as the ISDP data). The ISDP data base was a precursor of SIPP, and displays many of the same attributes that will make SIPP especially advantageous for research on turnover. As such, our research illustrates the potential uses of SIPP for research in this area.

THE DATA

A successful analysis of food stamp turnover requires data possessing several essential characteristics. To our knowledge, the ISDP is the

first data base to exhibit all of these desirable features simultaneously. These characteristics are as follows:

- The ISDP is nationally representative, and it yields enough observations (about 7,500 households) to permit meaningful analysis.

- The ISDP is longitudinal; retrospective interviews were conducted quarterly for a period of fifteen months.

- Data on income that a sample household receives from various sources (including the receipt of food stamps) is ascertained on a month-by-month basis. Furthermore, the exact timing of all changes in household composition between interviews is pinpointed. These data requirements are especially crucial to the analysis of turnover in the Food Stamp Program because many households enter and exit the program in the space of less than a year, and because program participation may be triggered by events such as a temporary drop in household income that could not be picked up by a survey that only measures a household's annual income.

- The data on income, household composition, and assets permit us to simulate eligibility for the program, using the eligibility criteria actually used in 1979.

- There is a wealth of explanatory variables, such as household and individual characteristics, types of income, and program participation, available to the analyst for both tabular and multivariate analyses of turnover.

We created an analysis file from the survey data using the RAMIS II data base management system. Storing and manipulating the ISDP in this fashion facilitated the creation of an analysis file, and in particular, enabled us to deal with several practical problems, such as linking house

holds across waves and maintaining a consistent longitudinal definition of the household unit.[1]

THE ANALYSIS

We conducted analyses of turnover in both participation and eligibility to participate in the program. We focused on three measures of program turnover:

- The entry rate; that is, the probability that a household that was not receiving food stamps in a given month received food stamps the next month.
- The exit rate; that is, the probability that a household that received food stamps in a given month did not receive food stamps in the next month.
- The annual/monthly ratio; that is, the ratio of the probability that a household received food stamps over the course of a year to the probability that it receives food stamps in a given month.[2]

Both tabular and multivariate techniques were used in the analysis. The multivariate analysis used the RATE model for the analysis of event histories that has been developed by Nancy Tuma and colleagues (Tuma et al., 1979), and applied in studies of marital instability, unemployment duration, and other socioeconomic phenomena. Although the tabular analysis

---

[1]One element of the ISDP survey design, which is also being implemented in SIPP, is that individual household members are followed and interviewed after leaving the sample household members are followed and inteviewed after leaving the sample household. This sometimes creates ambiguity as to which sample household at a given point in time is the correct "successor" to a sample household as of an earlier point. Our procedure for resolving this ambiguity is described in a longer version of this paper.

[2]The measures of turnover in eligibility, as opposed to participation, are defined in an analogous fashion.

of exit and entrance rates focused on one characteristic at a time, making it difficult to sort out compositional effects, the multivariate analysis in general yielded largely similar results.

Our principal findings were as follows:

Turnover in Participation

- Of all households who received food stamps in a given month, about 7.3 percent had exited from the program within the next month.
- Given that a household did not receive food stamps in a given month, there was a 0.5 percent probability that it entered the program in the next month.
- The probability that a household was a food stamp recipient household at least once in the course of the year was about 1.7 times the probability that it participated in the program in a given month.
- Among socioeconomic groups of interest, both the tabular and multivariate analyses indicated that the lowest monthly exit probabilities were exhibited by households that received AFDC and/or other types of welfare, nonwhite households, households containing disabled persons, households in which no person is employed, and households headed by a single person.
- The highest probabilities of entrance into the program were exhibited by households that received AFDC, households headed by a single person with children, nonwhite households, large households, and households in which no person is currently employed.

Turnover in Eligibility

- There appears to be substantial turnover in food stamp eligibility. In our sample, the probability that an eligible household became ineligible each month was about 17 percent, and the probability that a previoussly ineligible household became eligible was about 6 percent.

- The types of households that have the highest propensity to become eligible for Food Stamps, as indicated by both tabular and multivariate estimates, are those that receive AFDC and other types of welfare, households headed by a single person with children, nonwhite households, and households containing elderly or disabled persons.

- The types of households that have the lowest propensity to leave eligible status are those that receive AFDC and other types of welfare, households with a single head, households in which no one is employed, and household containing an elderly or disabled person.

SUMMARY

We believe that our research, in addition to meeting a current need for policy-relevant research on the Food Stamp Program, illustrates some of the vistas for new research that will be opened up by SIPP.[3] In the specific area of turnover in the Food Stamp Program, we see several potentially fruitful avenues for future research. Indeed, we are pursuing some of these areas ourselves in ongoing work sponsored by the U.S Food and Nutrition Service using the ISDP data. These include:

- Techniques for performing longitudinal imputations of data in cases of nonresponse should be developed and implemented. This move would materially improve the accuracy with which we simulate eligibility for the Food Stamp Program.

- The interrelationships among patterns of participation in various income maintenance programs should be studied at greater length.

---

[3]Indeed, given that the SIPP data will not be available for public use for some time, as well as the fact that the research community has barely scratched the surface, so as to speak, in terms of the potential uses of the ISDP data we would hope that our research reported here will stimulate further ISDP-based research.

- The relation between events such as Food Stamp entry and exit and changes in household characteristics should be investigated. The current study has focused largely on a household's status (e.g., being a household with just one head) rather than the impact of changes in status (e.g., a recent separation or divorce).

- The question of possible labor supply effects of the Food Stamp Program and their implications for observed levels of participation and turnover should be explored.

- Further research should be performed on such phenomena as duration dependence in turnover. This is especially important in connection with predicting the number of long-term participants. Because the restrictive statistical assumptions used throughout most of this report may not be a good approximation to reality, it is difficult to predict the average expected duration of participation from the one-year sample.

The Measurement of Household Wealth in the Survey of Income and Program Participation

by

Enrique J. Lamas
and
John M. McNeil

Population Division
U.S. Bureau of the Census

I. Introduction

The Survey of Income and Program Participation (SIPP) is designed to obtain current estimates of income, labor force activity, and participation in government transfer programs. Currently available survey data have important limitations which have been well recognized and documented [Ycas, 1982; Ycas and Lininger, 1982; Nelson, McMillen, and Kasprzyk, 1983]. The March supplement to the Current Population Survey (CPS) is presently the major source data on the distribution of income. In the March CPS, household members at the time of the interview are asked to recall their income for the previous calendar year. While the CPS survey does well in estimating wage and salary income, it has serious underreporting problems with respect to property income (such as interest, dividend, and rental income) and several other income types (such as Supplemental Security Income (SSI), and worker's compensation)[1]. The CPS has other limitations: subannual income estimates are not available; annual household income estimates cannot be adjusted for changes in household composition or size during the year, for example, as a result of a death, marriage, or divorce; and, information important for assessing economic well-being of the population and for policy analysis, such as assets, taxes and other characteristics, are not covered.

SIPP is designed to overcome many of these limitations. Income data from a wide variety of sources including wage and salary and government transfer programs are collected on a monthly basis. Changes in household composition are also identified on a monthly basis. In addition to improving measures of income and program participation, SIPP will also obtain detailed data on other important topics including asset and liabilities, and taxes.

---

[1] When comparing CPS income aggregates to independent benchmark estimates, CPS captures 97.4 percent of independent wage and salary totals. However, the CPS only covers 41.5 and 44.1 percent of interest and dividend totals, 69.4 percent of SSI totals, and 42.3 percent of worker's compensation totals.

Asset and liability data are important in determining program eligibility and assessing the economic situation of families. Information on the distribution of household wealth is important and changes in wealth provide data on consumer savings. SIPP will be unique among surveys in providing a recurring series on household wealth. The focus of this paper is to describe the effort of SIPP to provide estimates of household wealth holdings in the United States and to present some preliminary results.

Data to study the composition and distribution of household and personal wealth have come from three sources: estate tax returns, synthetic databases, and surveys [Wolff, 1983]. Each source, however, has its limitations. Estate tax returns consist of records filed with the Internal Revenue Service (IRS) for estate tax assessments. Coverage of the population is a major problem for this data source. Only descendents with substantial wealth holdings ($300,000 or more in 1981 and $500,000 or more in 1982) are required to file estate returns. Under certain assumptions about mortality, an "estate multiplier technique" has been used to estimate wealth for living individuals who would have been required to file estate tax returns [Schwartz, 1983]. However, this technique has limited population coverage, and only provides estimates for top wealth holders.

Synthetic databases, such as the Measurement of Economic and Social Performance file (MESP) [Wolff, 1983] and the 1973 Office of Tax Analysis file [Greenwood, 1983], merge data from several sources in order to get appropriate population and asset coverage. The MESP file consists of a statistical match of the 1970 census 1/1,000 public use sample to IRS tax returns. All asset values were imputed from the IRS tax information for financial

assets and from the Consumer Expenditure Survey (CES) for consumer durables. The Office of Tax Analysis file consists of a statistical match of the 1973 CPS to tax records from the 1973 Individual Income Tax Model. The IRS data on dividends and interest are used to estimate the market value of financial assets and the IRS data on property taxes are used to estimate the value of real estate holdings. Net wealth estimates are then imputed using a regression of net wealth on asset values based on IRS estate tax data. Limiting these databases are various assumptions which underlie the matching procedures, the procedure of income capitalization, and the extension of estimates to the whole population [Smith, 1983].

A third source of data are surveys covering household assets and liabilities. Major previous surveys include the Survey of Consumer Finances (SCF), the Survey of Financial Characteristics of Consumers (SFCC), and the Income Survey Development Program (ISDP). The SCF are periodic surveys (the latest conducted in 1983) with sample sizes of 3,500 to 4,000 consumer units. The SFCC, conducted in 1962 and 1963, canvassed 2,557 units. Information on the size and components of wealth was collected as of December 31, 1962 [Projector and Weiss, 1966]. Both the SFCC and the 1983 SCF used IRS records in order to oversample high income households. The ISDP, conducted in 1979 and 1980, collected information on assets and liabilities from approximately 7,000 households. The ISDP was a research panel designed to prepare for SIPP. The ISDP interviewed respondents on a quarterly basis on six occasions. In the fifth interview, a supplement on assets and liabilities was administered which collected information as of December 31, 1979.

Surveys have the advantage of being able to have samples and survey instruments specifically designed to gather the necessary detail and information

to estimate wealth holdings. Surveys, however, have suffered from limitations in population and asset coverage.

In the next section of this paper, significant features of the SIPP design with respect to the measurement of wealth are discussed. The final section presents selected results from the ISDP, the pretest of the SIPP household wealth module, and the first wave of SIPP.

II. SIPP Design Features

SIPP is a panel survey consisting of approximately 20,000 households which are interviewed every four months for a period of 2 1/2 years. As SIPP progresses, new panels will be started every year which will allow cross-sectional analyses based on a total sample of approximately 35,000 households. At each interview, information on income, program participation, and other characteristics for each of the previous four months is obtained for each person. Persons who move during the life of the panel, are followed and interviewed at their new addresses.[2] Questionnaire items include a "core" set of questions which are repeated in each wave of interviewing. These items cover labor force participation, detailed income recipiency, and participation in government programs. For waves 2 through 8, the core items are updated and the questionnaire is expanded with additional questions on items not covered in the core. Detailed questions concerning the amounts of personal and household asset and liabilities are included in wave 4 which is to be conducted in September through December 1984. These items will be updated one year later in wave 7. A list of the detailed asset and liability types covered is presented in Appendix A. Asset and liability coverage is comprehensive.

---

[2] Persons who move individually or in groups are followed if they relocate within 100 miles of any SIPP PSU. Persons who move into institutions are not interviewed if they are institutionalized.

The SIPP design is expected to have a positive effect on the ability to measure wealth. Research has found that a major source of bias in survey estimates of wealth is the nonreporting of asset ownership [Ferber, 1982]. Additionally, there is some nonreporting of asset values. Several features of SIPP are expected to have positive effects on the reporting of ownership and value of assets [Radner and Vaughan, 1984].

First, asset ownership questions precede questions on asset values. Separating ownership and amount questions helps focus on identification of asset holdings. In addition, the relatively sensitive amount questions do not negatively impact on the reporting of asset ownership.

Second, the longitudinal nature of SIPP helps identify asset ownership. Asset ownership information is collected in each wave. In the initial interview, a set of detailed questions designed to identify ownership of income earning assets are asked for each person in the household. An asset roster is created and recorded in the control card. In subsequent interviews, the respondent's asset roster for the previous wave is pretranscribed to the current questionnaire. During an interview, the pretranscribed asset roster is checked for accuracy. Then questions are asked to determine whether any assets have been liquidated or whether any new ones have been acquired. With this procedure, relatively accurate asset ownership information is obtained before respondents are asked about asset values and amounts of liabilities in wave 4.

In longitudinal surveys, attrition, that is dropping out of the sample is of concern. While some respondents leave the sample, there is evidence to suggest that cooperation or rapport obtained in repeated interviews increase the reliability of financial data [Ferber and Frankel, 1978]. Furthermore, the longitudinal nature also provides the opportunity to gather information

missed in a previous interview. If one interview is missed for the household or an individual respondent, a "Missing Wave" section is completed. In this section, a limited set of key questions are asked concerning labor force participation, income recipiency, and asset ownership during the missed wave.

Third, ownership, income, and value of assets are asked for each individual by type of ownership. Information is gathered for assets held jointly with spouse and for assets held individually. There is some evidence that nonreporting of ownership is higher for assets held by one individual as compared to assets held jointly with others [Ferber and Frankel, 1978]. Asking respondents by type of ownership directly may tend to reduce differential nonresponse between assets held in own name and jointly with others. In addition, collecting income and asset values by type of ownership rather than one total amount for an asset may tend to give more accurate income and asset value amounts. Because both asset income and asset value are collected, information about asset income can be used for assessing the reasonableness of the asset value data and for imputing missing values.

Finally, "callback items" have been introduced for critical questions concerning asset values. Callback items are designed to reduce nonreporting of income and asset amounts. For selected items, when a respondent answers "don't know," the interviewer reads a statement on the importance of the information requested and the respondent is asked whether it would be possible to call back later for an estimate of the amount. For respondents agreeing to be called back, a "reminder card" is provided to the respondent with the requested information checked. The interviewer telephones the respondent at an agreed upon time to obtain the missing information which is entered in a

section of the questionnaire reserved for callback amounts. The impact of the callback system is to reduce the incidence of missing information. Since callbacks involve added respondent and interviewer burden, only a limited number of items can be identified as callback items. To supplement the callback procedure, special instructions have been included for other important questions which instruct interviewers to probe for an estimate before accepting a "don't know" response. In this way, interviewers are alerted to key items for which they should give special effort to obtain estimates.

Many of these features were tested in the 1979 ISDP research panel which collected asset and liability information in wave 5[3]. In the ISDP, asset ownership questions preceded income value of asset items, asset rosters were updated each wave, and income and value of assets were obtained at the person level by type of ownership (jointly with spouse, in own name, and jointly with others). However, there are some differences between the surveys. The detailed asset types covered by ISDP and SIPP differ as the result of the availability of new types of assets since 1979. In addition, callback items were not utilized for asset items in the ISDP and special instructions were not used for important items. In any case, the ISDP should give some indication of results expected for SIPP. The next section presents some results based on the ISDP asset and liability information and some preliminary results from SIPP.

An important issue for surveys which measure wealth is population coverage. Evidence suggests that wealth holdings are concentrated. Studies have estimated that the top 5 percent of household hold approximately 30 to 50 percent of net wealth [Wolff, 1983; Greenwood, 1983; and Smith, 1983]. In addition,

---

[3] Home equity, automobiles, and life insurance information were collected in wave 2, primarily to evaluate government program eligibility.

holdings of certain assets such as stocks are even more highly concentrated with the top 5 percent holding approximately 70 percent of this asset. As a result, the normal SIPP area frame sample has a limited coverage of the top wealthholders. This problem was noted by the Census Advisory Committee on Population Statistics which recommended that "the Bureau explore the opportunity to augment the SIPP sample design every 5 years to oversample the upper end of the income and wealth distribution, where special effort would be directed to producing a reliable indication of the entire distribution of wealth" [U.S. Bureau of the Census, 1983].

The Census Bureau has done some preliminary work to explore the methods which might be used to obtain data files in which top wealthholders are adequately represented. One possibility is to use IRS data files to develop a list frame sample of top income holders. An alternative approach is to use estate tax data to estimate wealth distribution of top wealthholders. Using estate tax data in conjunction with SIPP data has the potential to improve population coverage. At this time, however, the work in this area is very preliminary. In any case, the probable limitations of SIPP wealth estimates and the need to improve population coverage have been recognized by the Census Bureau and the work to improve the estimates continues.

III. Data and Results

The 1979 ISDP research panel was designed to develop and test survey procedures and instruments to be implemented in SIPP. The ISDP was a panel survey where each household was interviewed at quarterly intervals on six occasions. The fifth wave, which took place between January and March 1980, asked questions on asset and liability amounts as of December 31, 1979. Approximately 7,000 households were interviewed.

Research has found that ISDP income data has substantial improvement in coverage compared to CPS income data [Vaughan, Whiteman, and Lininger, 1983; Vaughan, 1983]. Income recipiency estimates from ISDP are higher than CPS figures for property and government transfer incomes, such as, interest, dividends, SSI, veteran's payments, and pensions (private, state, and local government). Income aggregates and benchmark comparisons also indicate greater coverage in ISDP as compared to CPS, especially in interest, dividends, SSI, unemployment compensation, alimony or child support, and pensions [Vaughan, 1983].

The ISDP wealth data showed some mixed results. Asset ownership information showed increases in identification of ownership as compared to other survey estimates. However, some asset values suffered from high nonresponse rates. While nonresponse rates for home and automobile equity and unsecured liabilities were generally low, the rates for value of stocks and mutual funds, and own business were generally high. The remainder of this paper, reports asset ownership results from wave 1 of SIPP and asset value results from SIPP, ISDP, and a pretest of the wave 4 wealth supplement. In particular, response rates for asset ownership and asset amounts are presented and patterns of asset ownership are analyzed.

A. Nonresponse Rates for Asset Ownership

Ownership of assets is established in the first wave of SIPP and an asset roster is constructed for each respondent. This roster is verified and updated in subsequent waves. The amounts of assets held are not systematically covered until the fourth wave topical module. In Table 1, the level of missing information on asset ownership are shown. Respondents can answer that they "don't know" if they own a specific asset type or they can refuse to answer. The nonresponse rate

Table 1.--Percent Missing Information For Asset Ownership By Type of Respondent, SIPP Wave 1

Asset type	Total			Proxy			Self			Unweighted base
	Total	Don't know	Refusal	Total	Don't know	Refusal	Total	Don't know	Refusal	
Total	1.4	0.4	1.0	1.9	1.0	0.9	1.2	0.2	1.0	40,959
Savings Accounts	1.7	0.4	1.3	2.1	0.9	1.2	1.4	0.1	1.3	27,124
Money Market Deposit Accounts ...	2.1	0.7	1.4	2.8	1.6	1.2	1.7	0.3	1.4	27,039
Certificates of Deposit	2.2	0.8	1.4	3.0	1.7	1.3	1.8	0.3	1.5	27,102
Interest Earning Checking Accounts	1.9	0.6	1.3	2.8	1.5	1.3	1.6	0.3	1.3	27,098
Other Interest Earning Assets[1]	1.4	0.4	1.0	1.7	0.9	0.8	1.2	0.2	1.0	40,858
Stocks and Mutual Funds	1.2	0.4	0.8	1.6	0.8	0.8	1.1	0.2	0.9	40,971
Rental Property	0.9	0.1	0.8	1.0	0.3	0.7	0.9	0.1	0.8	40,959
Royalties	0.9	0.2	0.7	1.1	0.4	0.7	0.9	0.1	0.8	40,928
Other Investments	1.3	0.5	0.8	1.6	0.9	0.7	1.1	0.3	0.8	40,935

[1]Consist of money market funds, U.S. Government securities, municipal and corporate bonds, mortgages, and savings bonds.

for all asset types is low at 1.4 percent of all persons asked about asset ownership. For specific asset types, the rates differ and range from 0.9 percent for rental property and royalties to 2.2 percent for certificates of deposit[4]. When the nonresponse rates are decomposed, the refusal rates are generally higher than the "don't know" rates. This result is true of self respondents. In general, self respondents have higher refusal than "don't know" rates. However, as would be expected, the frequency of "don't know" responses for asset ownership is greater for proxy respondents which range from 0.3 to 1.7 percent, as compared to self respondents which range from 0.1 to 0.3 percent. The refusal rates for both types of respondents were similar (ranging from 0.7 to 1.5 percent). As a result of higher "don't know" rates, proxy respondents had somewhat higher overall nonresponse rates. For both types of respondents, however, the absolute level of the nonresponse rates are low.

To examine the asset ownership information further, nonresponse rates by demographic and socioeconomic characteristics are shown in Table 2. Several patterns emerge. Nonresponse rates for each asset type are approximately the same between sex or race groups. Some of the rates, however, are significantly different by age and education levels. The nonresponse rates for each asset type increase with the age of the respondent. For respondents in the less than 25 and 25 to 34 age groups, nonresponse rates range between 0.3 to 1.2 percent; for the 35 to 44 age group, rates range from 0.9 to 2.0; for the

---

[4]Significance tests were performed on the differences between rates using the formula

$$\frac{(P_1-P_2)}{\sqrt{\frac{P_1(1-P_1)}{N_1} * F_1 + \frac{P_2(1-P_2)}{N_2} * F_2}}$$

where P_1 and P_2 are the proportion of nonresponses, F_1 and F_2 are sample design factors, and N_1 and N_2 are the number of sample cases for each proportion. Differences noted in the text are significant at the 95 percent confidence level.

Table 2.--Asset Ownership Nonresponse Rates by Characteristic, Wave 1 SIPP

Characteristics of reference person	Asset type										
	Total	Savings accounts	Money market deposit accounts	Certificates of deposit	Interest earning checking accounts	Other interest earning assets	Stocks or mutual funds	Rental property	Royalties	Other financial investments	Unweighted base
Total	1.4	1.6	2.1	2.2	2.0	1.4	1.2	0.9	0.9	1.3	40,797
AGE											
Less than 25 years ...	0.6	0.7	0.9	0.9	1.0	0.5	0.4	0.3	0.3	0.6	8,981
25 to 34 years	0.7	0.8	1.0	1.1	1.2	0.7	0.7	0.4	0.5	0.8	8,790
35 to 44 years	1.4	1.5	2.0	2.0	1.9	1.4	1.4	1.0	0.9	1.5	6,610
45 to 54 years	1.9	2.4	2.7	2.9	2.5	1.8	1.6	1.1	1.1	1.5	5,264
55 to 64 years	2.0	2.3	2.8	2.8	2.5	2.1	1.8	1.3	1.4	1.8	5,046
65 and over	2.5	2.7	3.4	3.5	3.0	2.6	2.1	1.7	1.7	2.3	4,715
RACE											
White	1.5	1.6	2.1	2.2	1.9	1.4	1.3	0.9	0.9	1.3	35,365
Black	1.2	1.9	2.3	2.5	2.6	1.1	0.9	0.6	0.6	1.1	4,347
Other	0.7	0.9	1.4	1.4	0.9	0.7	0.4	0.3	0.3	0.6	1,085
SEX											
Male	1.5	1.7	2.2	2.3	2.1	1.5	1.3	1.0	1.0	1.4	19,165
Female	1.3	1.6	2.0	2.0	1.8	1.3	1.2	0.8	0.8	1.2	21,632
EDUCATION LEVEL											
Less than high school	1.0	1.5	1.8	1.7	1.6	0.9	0.7	0.6	0.6	0.9	12,540
High School	1.6	1.8	2.2	2.3	2.1	1.5	1.3	1.0	1.0	1.4	13,244
Some college	0.9	1.0	1.3	1.4	1.2	1.0	0.8	0.5	0.5	0.5	8,270
College graduate	2.0	1.8	2.4	2.6	2.2	2.1	2.1	1.4	1.3	1.9	6,347

45 to 54 and 54 to 64 age groups, rates range from 1.1 to 2.8; and for the 65 and older age group, rates range from 1.3 to 3.5 percent. While the rates for the oldest age group are significantly higher than for the youngest group for each asset type, it should be noted that even the highest rates (3.4 and 3.5 percent for money market deposit accounts and certificates of deposit, respectively) are relatively low.

Nonresponse rates also differ by education level. The nonresponse rates for ownership of several asset types is higher for college graduates than for respondents who did not complete high school or who completed some college. Nonresponse rates for college graduates range from 1.3 to 2.6 percent, while the rates for respondents with less than high school or some college education levels range from 0.5 to 1.8 percent. Differences in rates between these groups are significant for each detailed asset types, except saving accounts and interest earning checking accounts.

B. Asset Ownership Patterns

Analyzing the SIPP asset ownership results, several patterns emerge. The frequency of asset ownership is shown in Table 3. The most frequently held assets were savings accounts and home ownership with 65.7 and 64.1 percent of households reporting ownership, respectively. The ownership of the remaining assets was reported by approximately 20 percent of households, with rental properties and royalties reported by 10 percent or less of households. Of interest are the asset types which became newly available after 1982 as a result

Table 3.--Percent of Households Reporting Ownership of Asset Types[1]

Asset type	SIPP Wave 1
Savings accounts	65.7
Money market deposit accounts	17.1
Certificates of deposit	19.0
Interest earning checking accounts	23.1
Other interest earning assets[2]	18.1
Stocks or mutual fund shares	20.9
Rental property	10.2
Royalties	1.4
Home ownership	64.1

[1] Preliminary SIPP wave 1 results.

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

of deregulation in the banking industry. Approximately 23 percent of the households reported ownership of one or more interest earning checking accounts, while 17 percent of the households reported ownership of money market deposit accounts.

The SIPP wave 1 results for selected assets are compared to asset ownership data from other surveys in Table 4. In general, the ownership estimates obtained in wave 1 of SIPP are similar to estimates derived from other sources. The only major difference occurs in savings accounts. The Consumer Credit Survey and the ISDP results show approximately 75 percent of households reporting ownership of savings accounts, while SIPP results show approximately 66 percent of the households owning such accounts. A plausible explanation is that savings accounts are highly substitutable with the newly available assets. Individuals have an incentive to switch from savings accounts to new accounts for the greater liquidity available with interest earning checking accounts and for the higher interest rates available with money market deposit accounts. The result is a negative impact on the percent of households owning savings accounts.

C. Asset Amount Response

An important feature of the SIPP design is to have income recipiency and asset ownership questions precede income and assest amounts questions. Once income recipiency is established for a respondent, questions on amount of income are asked by income type. Income from assets, e.g. interest, dividends and rental income, are covered in each wave. This section focuses on the reporting of property income and asset amounts in the first wave of SIPP.

The questions on asset income are divided by type of ownership, that is, assets held jointly with spouse, and assets in own name or held jointly with

Table 4.--Percent of Households Reporting Ownership of Selected Assets

Asset Type	SIPP wave 1 (1983)	ISDP (1979)	Consumer Credit Survey (1977)	Consumer Expenditure Survey (1972-73)	Survey of Financial Characteristics of Consumers (1962-63)
Savings accounts	65.7	73.8	76.9	58.9	59.0
Certificates of Deposit	19.0	15.7	13.6	-	-
Stocks or Mutual Fund Shares	20.9	19.6	24.7	-	16.0
Rental Property	10.2	13.6	-	-	-
Home ownership	64.1	60.5	61.0	-	57.0
Sample size	19,878	6,992	2,563	9,914	2,557

others. Persons identified as holding an asset are asked questions concerning the amount of income received. If respondents cannot provide the amount of interest earned, the average balances of accounts are asked in order to impute the interest income received. A callback is provided for the balance amount if the person does not know the amount at the time of the interview but can provide an estimate later. Unlike the ownership items, questions on interest amounts refer to grouped assets. In particular, assets held with financial institutions (savings accounts, money market deposit accounts, certificates of deposit, and interest earning checking accounts) are grouped together and the total interest earned is asked. Simflarily, other interest earning assets (money market funds, U.S. Government securities, municipal or corporate bonds, and other interest earning assets) are grouped and interest income from those sources is covered.

The reporting of interest income is summarized in Table 5. On the average, two-thirds of respondents reported an amount of interest earned. Approximately 15 to 20 percent of respondents did not know the interest income amount, but provided an estimate of the total balance in the accounts. In general, over 80 percent of respondents gave the amount of interest or the balance in the accounts. Less than 11 percent did not know the amount of interest and did not provide a balance amount. Only 5 to 8 percent of respondents refused to report the interest income.

These patterns of reporting did not differ by type of ownership, that is, joint with spouse versus in own name or with others. However, the frequency of persons who refused or did not provide an interest/balance amount was lower on average for assets at financial institutions (11.7 percent) than other interest earning assets (17.1 percent).

Table 5.--Distribution of Interest Income Reporting

Asset Type	Total	Interest income reported	Don't Know		Refusal	Unweighted base
			Asset amount reported	No asset amount reported		
TOTAL						
Assets at financial institutions ..	100.0	66.8	21.5	7.2	4.5	19,664
Other interest earning assets	100.0	67.6	15.3	10.2	6.9	1,701
JOINT WITH SPOUSE						
Assets at financial institutions ..	100.0	68.0	20.0	6.5	5.5	7,603
Other interest earning assets	100.0	60.4	13.6	10.8	8.1	664
OWNERSHIP IN OWN NAME						
Assets at financial institutions ..	100.0	66.0	22.5	7.6	3.9	12,061
Other interest earning assets	100.0	69.2	15.8	9.2	5.8	1,037

Of interest are the cases which did not know the amount of interest earned and which were asked the average balance held for imputation purposes. Over 5,900 respondents were asked the balance of accounts held with financial institutions, and 455 were asked the balance of other interest earning accounts. For assets held with financial institutions, approximately 78 percent of the respondents which reported they did not know the interest income earned were resolved using the average balance reported; for other interest earning assets, 67 percent of the "don't know" cases reported a balance amount. Only 6 percent of the respondents refused to report an amount.

The results presented above deal with the questions in the wave 1 of SIPP which asked for asset income amounts, specifically interest earning assets. For other assets, questions on the amount held in asset will be asked in the fourth wave supplement. A pretest of the fourth wave questionnaire was conducted in Atlanta, Georgia during February 1984. For the pretest, the sample was not scientifically selected and low income and high income households were oversampled. Therefore, caution should be exercised in drawing conclusions based on the pretest since the sample is not representative and the sample size is small. However, some pretest results should be pointed out. To give some indication of response rates for asset amount questions, nonresponse rates for selected items in the pretest and similar items in ISDP are shown in Table 6.[5]

For asset amounts (first section of Table 6), ISDP had relatively high nonresponse rates. The SIPP pretest results appeared to show some improvement. The strategy used in the wave 4 questionnaire to improve response rates (increase the number of callback items, identify critical items for which interviewers

---

[5] In the pretest, respondents were asked callback questions, but callbacks were not followed-up after interviews due to time constraints. In wave 1 of SIPP, approximately 66 percent of the respondents agreeing to the callback reported an amount. To simulate the effect of callbacks, two-thirds of respondents which agreed to a callback were counted as having provided response and were not counted as missing information.

Table 6.--Comparison of Nonresponse Rates on Amounts Between ISDP and SIPP Wave 4 Pretest

Asset Type	ISDP	SIPP Pretest (Net of Callback)
Amount in savings accounts	24.9	16.8
Amount in cash and checking accounts	23.1	11.5
Market value of stocks	59.6	21.0
Amount in bonds and government securities	32.2	25.9
Value of own business	42.8	22.8
Debt on own business	22.2	12.4
Unsecured Liabilities		
Credit card/store debts	10.1	12.8
Medical bills	8.7	6.9
Debts to private individuals	5.2	0.3
Debts to financial institutions	8.6	15.4
Current Employer Retirement Plan		
Employer has pension plan	8.9	0.0
Included in plan	2.7	2.7
Profit sharing plan	9.8	20.0
Employer makes payments	8.4	8.6
Employee makes payments	3.9	2.9
Employee has vested rights	14.2	15.2

should probe for estimates, and stress interviewer training) appeared to have provided positive results in the pretest sample. Pretest nonresponse rates for unsecured liabilities and retirement plan items were generally close to the ISDP results. ISDP and pretest rates for these items tended to be lower than for the asset amounts.

In Table 7, some detailed nonresponse rates from the pretest are presented. Two factors should be noted. First, refusal rates are low. Second, the callback items appeared to work well. Comparison of the "don't know" and "don't know net of callbacks" columns indicate that callbacks reduce the nonresponse rates.

IV. Conclusion

The SIPP will provide annual estimates of wealth. Information on asset and liabilities is useful for many types of analyses, including program eligibility simulation studies and measurements of the distribution of wealth. In this paper, significant features of the survey were presented along with some preliminary results from SIPP. Several design features of SIPP, including longitudinal nature of the survey design, separating asset ownership and asset value questions, updating asset roster each interview, callbacks, probe instructions, and a missing wave section, are likely to have a positive effect on the reporting of asset ownership. In addition, asset and liability coverage is comprehensive. Results from the first wave of SIPP show that nonresponse rates for asset ownership are low. In addition, frequency of ownership patterns are reasonable and the results are comparable to findings from other surveys.

Table 7.--Nonresponse Rates for Wave 4 Pretest

Item	Don't know	Don't know net of callbacks	Refused	Unweighted base
Savings Accounts, Money Market Deposit Accounts, Certificates of Deposit and NOW Accounts				
Total balance for accounts joint with spouse	16.1	5.3	0.0	31
Total balance of accounts in own name	26.8	23.1	3.6	56
Stocks and Mutual Fund Shares				
Market value of stocks in own name	27.5	22.5	0.0	40
Own Home				
Have mortgage on home	0.0	-	0.0	67
Number of mortgages on home	0.0	-	0.0	36
Amount of mortgage owed	19.4	-	2.8	36
Current market value of home ...	19.1	-	0.0	68
Year home purchased	2.9	-	0.0	68
Purchase price	17.6	-	1.5	68
Amount of real estate taxes paid	23.4	-	0.0	64
Amount of cash on hand and in checking accounts	10.2	-	1.3	236
Have an educational loan	0.8	-	0.0	240

- Not applicable

Preliminary results also show some indication of improvements in nonresponse rates for items on asset amounts as compared to the relatively high nonresponse rates for questions on asset values found in ISDP.

REFERENCES

[1] David, Martin H., "Technical, Conceptual, and Administrative Lessons of the Income Survey Development Program (ISDP)-Papers Presented at a Conference," Social Science Research Council, Washington, D.C., 1982.

[2] Ferber, Robert, John Forsythe, Harold Guthrie, and E. Scott Maynes, "Validation of Consumer Financial Characteristics: Common Stock," Journal of the American Statistical Association, Volume 64, Number 326, June 1969, pp. 415-432.

[3] Ferber, Robert and Matilda Frankel, "Evaluation of the Reliability of the Net Worth Data in the 1979 Panel: Asset Ownership on Wave 1," Survey Research Laboratory, University of Illinois, August 1981.

[4] Ferber, Robert and Matilda Frankel, "The Collection, Measurement, and Evaluation of Savings Account Reports," Survey Research Laboratory, University of Illinois, March 1978.

[5] Greenwood, Daphne, "An Estimation of U.S. Family Wealth and Its Distribution From Microdata, 1973," The Review of Income and Wealth, Series 29, Number 1, March 1983.

[6] Nelson, Dawn, David McMillen, and Daniel Kasprzyk, "Overview of the Survey of Income and Program Participation," U.S. Bureau of the Census, December 1983.

[7] Pearl, Robert and Matilda Frankel, "Composition of the Personal Wealth of American Households at the Start of the Eighties," Survey Research Laboratory, University of Illinois, presented at the American Economic Association meetings, December 1981.

[8] Pearl, Robert, Matilda Frankel, and Richard Williams, "The Effects of Missing Information on the Reliability of Net Worth Data From the 1979 ISDP Research Panel," Survey Research Laboratory, University of Illinois, May 1982.

[9] Projector, Dorothy S. and Gertrude S. Weiss, "Survey of Financial Characteristics of Consumers," Board of Governors of the Federal Reserve System, August 1966.

[10] Radner, Daniel B. and Denton R. Vaughan, "The Joint Distribution of Wealth and Income For Age Groups, 1979," Social Security Administration, Office of Research and Statistics, Working Paper 33, March 1984.

[11] Schwartz, Marvin, "Trends in Personal Wealth, 1976-1981," Statistics of Income Bulletin, Volume 3, Number 1, Summer 1983.

[12] Smith, James D., "Recent Trends in the Distribution of Wealth: Data, Research Problems and Prospects," presented at C.V. Starr Center Conference on International Comparisons of the Distribution of Household Wealth, New York University, November 1983.

[13] Smith, James D. and Stanton Calvert, "Estimating the Wealth of Top Wealth-Holders From Estate Tax Returns," 1965 Proceedings of the Business and Economic Statistics Section, American Statistical Association, pp. 248-265.

[14] U.S. Bureau of the Census, "Minutes and Report of Committee Recommendations, Census Advisory Committee on Population Statistics," April 8, 1983.

[15] U.S. Bureau of the Census, "Money Income of Families and Persons in the United States: 1980," Current Population Reports, Series P-60, No. 132, U.S. Government Printing Office, Washington, D.C., 1982.

[16] Vaughan, Denton R., T. Cameron Whiteman, and Charles A. Lininger, "The Quality of Income and Program Data in the 1979 ISDP Research Panel: Some Preliminary Findings," presented at the Social Science Workshop on the Survey of Income and Program Participation, Baltimore, Maryland, December 1983.

[17] Whiteman, T. Cameron, "Comparisons of the Distribution of Household Wealth From the 1979 ISDP and 1962 SFCC," presented at the C.V. Starr Center Conference on International Comparisons of the Distribution of Household Wealth, New York University, November 1983.

[18] Wolff, Edward N., "The Size Distribution of Household Disposable Wealth in the United States," The Review of Income and Wealth, Series 29, Number 2, June 1983, pp. 125-146.

[19] Ycas, Martynas A. and Charles A. Lininger, "The Income Survey Development Program: Design Features and Initial Findings - Paper 1," presented at Social Science Research Council Conference, Washington, D.C., October 1982.

[20] Ycas, Martynas A., "Measuring Annual Income: Conceptualization - Paper 2," presented at Social Science Research Council Conference, Washington, D.C., October 1982.

APPENDIX A

List of Asset and Liability Information Collected By the Survey of Income and Program and Participation

Interest Earning Assets

Assets

1. Regular/passbook savings accounts
2. Money market deposit accounts
3. Certificates of deposit or other savings certificates
4. NOW, Super NOW or other interest earning checking accounts
5. Money market funds
6. U.S. Government securities
7. U.S. Savings Bonds (type E and EE)
8. Municipal or corporate bonds
9. Other interest-bearing assets

Information collected for each asset

A) For asset held jointly with spouse
   - Ownership of Asset (Y/N)
   - Interest earned during reference period
   - Balance as of the last day of the reference period
   - *Call back later (Y/N)

B) For assets held in own name
   - Ownership of Asset (Y/N)
   - Interest earned during reference period
   - Balance as of the last day of the reference period
   - *Call back later (Y/N)

* Asked only if asset amount is not reported.

Dividend Earnings Assets

Assets

1. Stocks and mutual funds

Information collected

A) For asset held jointly with spouse
   - Ownership of asset (Y/N)
   - Market value as of the last day of reference period
   - *Call back later (Y/N)
   - Dividend income received during reference period
   - Any debt against asset (Y/N)
   - Amount of debt

B) For asset held in own name
   - Ownership of asset (Y/N)
   - Market value as of the last day of reference period
   - *Call back later (Y/N)
   - Dividend income received during reference period
   - Any debt against asset (Y/N)
   - Amount of debt

IRA/KEOGH Accounts

Assets

1. IRA accounts

2. KEOGH accounts

Information collected for each asset type

- Ownership of asset (Y/N)
- Number of years contributed to accounts
- Balance in accounts as of the last day of the reference period
- *Call back later (Y/N)
- Types of assets held in account

* Ask only if asset amount is not reported.

Own Home

Assets

1. Own home

2. Mobile home

Information collected

A) For own home
- Home being owned or being bought (Y/N)
- Current market value of home
- Year home purchased
- Total purchase price excluding closing costs
- Mortgage, deed of trust or other debt on home (Y/N)
- Current amount owed on mortgage or debt
- *Original amount of mortgage or debt
- *Year mortgage taken out
- *Monthly payments excluding taxes and insurance
- *Interest rate of mortgage or debt
- Variable interest rate (Y/N)
- Mortgage obtained through lower cost mortgage program (Y/N)
- Second mortgage or other debt on home (Y/N)
- Current amount owed on mortgage or debts
- *Original amount of loan
- *Year loan taken out
- *Term of loan
- *Monthly payments on loan
- Variable interest rate (Y/N)
- Mortgage obtained through lower cost mortgage program (Y/N)
- Any other mortgages on loans (Y/N)
- Principal owed on all other mortgages or loans
- Amount of real estate taxes paid on home

B) For mobile home
- Mobile home or site owned or being bought (Y/N)
- Current market value of mobile home
- *Year home built
- *Length of mobile home
- Mortgage, loan or other debt on mobile home (Y/N)
- Current amount owed on mortgage or loan
- Monthly payments excluding taxes and insurance

* Used for imputation. Asked only if information is not reported.

Rental Property

Assets

1. Residential, commercial or other rental property owned

Information collected

A) For rental properties held jointly with spouse
   - Ownership of rental property (Y/N)
   - Type of property
   - Market value of property
   - *Call back later
   - Mortgage on property (Y/N)
   - Amount owed on property

B) For rental properties held in own name
   - Ownership of rental property (Y/N)
   - Type of property
   - Market value of property
   - *Call back later
   - Mortgage on property (Y/N)
   - Amount owed on property

C) For rental properties held jointly with others
   - Ownership of rental property (Y/N)
   - Type of property
   - Market value of property
   - *Call back later
   - Mortgage on property (Y/N)
   - Amount owed on property
   - Share of equity in property

Other Real Estate Property

Assets

1. Nonrental real estate (other than own home)

Information collected

Ownership of (Y/N)
Share of equity in property

* Asked only if asset amount is not reported.

Motor Vehicles

Assets

1. Licensed car or truck (excluding recreational vehicles and motorcycles)

Information collected

A) For asset held by any member of the household
   - Ownership of asset (Y/N)
   - Number of vehicles owned
   - Who owns the vehicles

B) For 3 vehicles owned
   - Year, make and model of vehicles
   - Money owed on vehicle (Y/N)
   - Amount owed

Recreational Vehicles

Assets

1. Motorcycle, boat, recreational vehicle, or other vehicle

Information collected

A) For asset held by any member of the household
   - Ownership of vehicle (Y/N)
   - Number of vehicles
   - Who owns the vehicles
   - Type of vehicle (motorcycle, boat, snowmobile, recreational vehicle, other)

B) For 2 vehicles
   - Market value in present condition
   - Money owed on vehicle (Y/N)
   - Amount owed

Unsecured Liabilities

Liabilities

1. Store or credit card bills

2. Doctor, dentist hospital or nursing home bills not covered by insurance

3. Debt to private individual not in household

4. Loans from banks, credit union or financial establishment (excluding mortgages, vehicle loans, money owed to brokers)

5. Any other unsecured loan

Information collected for each liability

A) For liabilities owed jointly with spouse
   Existence of debt (Y/N)
   Amount owed as of the last day of reference period

B) For liabilities owed in own name
   Existence of debt (Y/N)
   Amount owed as of the last day of reference period

Business/Farm Equity

Assets

1. Business equity

2. Farm equity

Information collected for each asset

Total value of business
*Call back later
Total amount of debt owned on business
*Call back later
Percent of business owned

* Asked only if information on loan amount is not reported.

Retirement Plans

Asset

1. Pension or retirement plan

Information collected

Employer or union have pension or retirement plan (Y/N)
Covered by plan (Y/N)
Time necessary to be covered by plan
Why not covered by plan
Covered by more than one retirement plan (Y/N)
Defined benefit or defined contribution plan
Make payments toward plan (Y/N)
Employer contribute to plan (Y/N)
Time covered by plan (yr/months)
Vested rights in plan (Y/N)
Necessary time in plan for vested rights
Number of persons employed at this location
Number of persons employed at all locations
Profit sharing plan
Thrift plan or tax defered income plans available (Y/N)
Lump-sum payments possible (Y/N)

Life Insurance Policies

Asset

1. Life insurance policies

Information collected

Have life insurance policy (Y/N)
Face value of life insurance policies

THE WEALTH AND INCOME OF AGED HOUSEHOLDS

Daniel B. Radner

Social Security Administration

Preliminary Draft

July 1984

Paper prepared for presentation at the American Statistical Association meeting, Philadelphia, August 13-16, 1984.

I. Introduction*

The economic status of the aged population has been a topic of interest to researchers for some time, and there has been particular interest in recent years. In this paper we are interested in the economic resources of aged households. This topic is examined using data on both income and wealth. Estimates are presented and compared for all aged households and for aged households with both low income and low wealth. Socioeconomic characteristics, mean and median amounts, and the composition of wealth and income are discussed.

Many different indicators of economic well-being have been used by researchers. Income, specifically money income before taxes, is the most frequently used measure of economic well-being. Money income before taxes is also the definition of income used here. Of course, this definition is far from an ideal measure of economic well-being. An important exclusion is noncash income, which is a significant source of economic resources for most groups of the population. Also, by using pre-tax income, the resources available to the unit can be distorted.

Measures of economic well-being that are confined to income omit the wealth of the unit, although income from assets ordinarily is included in income. A relatively new data base, the 1979 Income Survey Development Program (ISDP) file, makes the examination of both income and wealth possible. Wealth and income data can be used together in several different ways. In this paper a simple approach is used--income and wealth are used as a two-dimensional classification. This approach does

not spread the wealth out over the expected lifetime of the unit as some methods (e.g., Weisbrod and Hansen 1968) do, but is concerned with a much shorter time horizon. In this application, the liquidity of the assets held can be very important. Although the most comprehensive definition available (net worth) is used here as the measure of wealth, the composition of wealth is examined so that liquidity can be taken into account.

In a recent paper, Radner and Vaughan (1984) examined the joint distribution of wealth and income for various age groups and, using several different definitions, looked at the proportion of households with both relatively low wealth and relatively low income. They showed that, despite the fact that average net worth is fairly high for aged households, the joint distribution of income and wealth is such that the percentage of aged households (age 65 or older) with both relatively low income and relatively low net worth exceeds the percentage for each age group in the 25-64 age range.

Aged households identified using one of those definitions of low income and low wealth are examined further in this paper. The particular definition used is that the household had to be in the bottom 20 percent of all households ranked by size of income and in the bottom 40 percent of all households ranked by size of net worth. For both income and net worth the rankings were based on amounts that had been adjusted for size of household using an equivalence scale derived from the U.S. poverty lines.

All estimates shown in this paper are on a household basis. For convenience it is assumed that the income and wealth of the household are resources for all members of the household and only for members of the household. Thus, relatives of persons in the household who do not live in the household (e.g., parents or children) are assumed to have no claim on the resources of the household, and the household is assumed to have no claim on other resources. Most of the estimates shown in this paper take into account the number of persons in the household.

Because the characteristics and economic situations of aged households can differ substantially by age, estimates are shown for detailed age groups within the aged group. Space limitations and small sample sizes restrict the age detail that can be shown.

Section II briefly describes the data used and presents definitions of important concepts used in the paper. Section III provides a brief overview of the income and wealth of households in different age groups. In Section IV characteristics of all aged households and of aged households with both relatively low income and relatively low wealth are presented and discussed. A summary and conclusions are presented in Section V.

II. Data and Definitions

The data used in this paper are from the 1979 ISDP file (Radner and Vaughan 1984). The sample was nationally representative and both the low and high ends of the income distribution were oversampled slightly.

Detailed information on income, assets, debts, and socioeconomic characteristics was obtained. The data used in this paper are primarily from the fifth wave of this multi-wave panel. The fifth wave, which has most of the data on wealth, contains observations on about 6,900 households resulting from interviews in January, February, and March, 1980.

The ISDP income data suffer from the underreporting which is common to household surveys. However, overall the ISDP data appear to be better than the income data in the Current Population Survey (CPS). The quality of the wealth data is difficult to assess, although it is known that misreporting and nonresponse are substantial problems. While there is some evidence of marked increases in estimates of asset ownership compared to other surveys, unfortunately item nonresponse on asset values (for assets other than owner-occupied housing and vehicles) is quite high. Thus, very substantial proportions of the final asset value aggregates stem from values assigned using "hot deck" imputation. This is a problem that should be kept in mind in interpreting the estimates shown. The post-imputation estimates of net worth and of most asset types used in this paper suffer from substantial underreporting. Financial assets appear to show the highest percentage of underreporting. However, a rough comparison of survey aggregates to independent control aggregates suggests results that are similar to the 1962 Survey of Financial Characteristics of Consumers (Projector and Weiss 1966). It should be noted that the ISDP estimates of the extreme upper tail of the net worth distribution show a far lower share of net worth than is shown by other

sources of wealth data. This "absence" of the extreme upper tail results at least in part from an emphasis in the survey on obtaining data for low- and middle-income units.

In general the demographic concepts used in this paper are the same as or similar to those employed by the Bureau of the Census in its Annual Demographic Supplement to the CPS. 1/ The data presented pertain to the civilian noninstitutional population living in the 50 states and the District of Columbia at the dates of interview. All estimates are on a household basis. Age classifications are based on the age of the household reference person (householder) at his or her most recent birthday. The householder is the person (or one of the persons) in whose name the home is owned or rented.

Income is defined on a before-tax basis and is presented at annualized rates; that is, as the measured three-month value times four. All money income received by household members during the three months preceding the month of interview is covered, including one-time or lump sum payments such as life insurance proceeds and gifts. Total money income consists of the sum of earnings, property income, OASDI, SSI, pensions, and other income. Earnings includes wages and salaries, net income from farm self-employment, and "draw" from nonfarm self-employment. 2/ Property income includes interest, dividends, rent, royalties, and income from estates and trusts. OASDI consists of social security and railroad retirement income. SSI consists of Supplemental Security Income, including both state and federal amounts. Pensions are the sum of government and private pensions.

Government pensions consists of U.S. civil service and military retirement pensions and state and local government pensions. Private pensions includes employer and union pensions and annuities. Other income includes unemployment compensation, worker's compensation, veterans' compensation and pensions, public assistance, educational benefits, alimony, assistance from relatives and friends, and lump sum payments.

The wealth concept used is net worth. Net worth consists of all assets less all debts covered by the survey. Net worth is defined to be wealth minus unsecured debt. All net worth components except home and vehicle equity are valued as of the end of 1979; those two components are valued as of mid-1979. 3/ Wealth is defined as the value of all assets covered by the survey less any debts secured by those assets. Several items sometimes included in estimates of wealth are not covered; for example, social security wealth, pension wealth, and trusts are excluded.

Wealth is the sum of the following items: home equity (owner-occupied); durable goods (equity in vehicles plus market value of household durables); business equity (nonfarm and farm); liquid financial assets (cash, checking accounts, passbook savings accounts, U.S. savings bonds); nonliquid financial assets (bonds, CD's, stocks, mutual fund shares); other assets (e.g., equity in other property, equity in nonactive business interest). Unsecured debt includes installment and noninstallment debt, unpaid medical bills, and educational loans.

III. General Patterns of Income and Wealth by Age of Householder

In this section an overview of the relationship between age and income and wealth is presented. Mean and median amounts are shown, and the dispersion of the income and wealth of aged households is discussed. Mean and median total money income and net worth are shown for age (of householder) groups in Table 1. 4/ Looking at income, these estimates show the familiar pattern of relatively low mean and median at young and old ages, with a peak in the 45-54 age group. The mean (median) for the 65 and over group is below the mean (median) for the youngest age group. Of course, the median is below the mean for every age group. The pattern for net worth is somewhat different. Mean amounts are low for the younger age groups, peak in the 55-64 group, and then decline. However, the mean (median) for the 65 and over group is far above the mean (median) for the three youngest age groups. In general, the median is about one half the mean in each age group. Within the detailed aged groups, both income and net worth decline as age increases, using both means and medians, with the 75 and over group substantially below the 65-69 group.

One way of looking at the mean amounts is to compute relative means (Table 2). The relative mean is the mean for the group divided by the mean for all households. Thus, the relative mean income for households age 65 and over is $13,060/$21,240 = 0.62. The first two columns of Table 2 show the relative means correponding to the means in Table 1. We see that although households age 65 or over have a relative mean income

that is only 62 percent of the overall mean, their relative mean net worth is 27 percent above the overall mean.

It is well-known that households with aged heads are, on average, smaller than households with nonaged heads. These differences in household size are adjusted for in this paper using an equivalence scale based on the U.S. poverty lines (U.S. Bureau of the Census 1981b, Table A-3).5/ A one-person household is taken as the base, and the income and net worth of each household are divided by the appropriate scale value to obtain amounts adjusted for size of household. 6/ The third and fourth columns of Table 2 contain the relative means obtained using the adjusted amounts. 7/ The 35-54 age groups, which have relatively large households, show declines from the unadjusted estimates. The 55-64 and 65 and over groups, which have relatively small households, show increases. The relative mean income for the 65 and over group rises to 0.79, while the relative mean net worth increases to 1.59. In general, the relative means of aged households rise about 25 percent. Thus, after the adjustment for household size, aged households appear to be somewhat better off relative to nonaged households than before the adjustment.

Table 3 shows the distribution of households with householder 65 and over and 75 and over among overall income quintiles and overall net worth quintiles. This table provides some information about the dispersion present in income and net worth amounts for those households. These estimates have also been adjusted for size of household in the manner described above. In the distribution by size of income, substantial

dispersion among the quintiles is present for both the age groups. For example, even though 59 percent of households in the 65 and over group are in the bottom two income quintiles, 11 percent appear in the top quintile. The distribution for the 65 and over group is somewhat higher than the distribution for the 75 and over group. The distributions among net worth quintiles also show substantial dispersion. There is some difference between the 65 and over and 75 and over distributions, with the distribution for the older group somewhat lower. Thus, this table shows that substantial dispersion in income and net worth amounts exists within these age groups.

IV. Characteristics of Aged Households

In this section all aged households and aged households with both relatively low income and relatively low net worth are discussed and compared. A household is defined to have relatively low income if it is in the bottom quintile (20%) in the overall (all ages) distribution of households by size of income, after adjustment for size of household. A household is defined to have relatively low net worth if it is in the bottom two quintiles (40%) in the overall (all ages) distribution of households by size of net worth, after adjustment for size of household. 8/ Households that have both low income and low net worth using these definitions will be called "LILNW" (Low Income and Low Net Worth) households.

Thirteen percent of all households are in the LILNW group. The percentages range from 9 percent for households age 55-64 to 21 percent

for households under 25. Fifteen percent of all households 65 and over are in the LILNW group; the percentage is 11 in the 65-69 group, 15 in the 70-74 group, and 19 in the 75 and over group. 9/ Thus, substantial percentages of the aged groups have both relatively low income and relatively low net worth. Table 3 shows that 24 percent of households age 65 and over are in the bottom two net worth quintiles, while 32 percent are in the bottom income quintile. Since only 15 percent are in both groups, it can be seen that about two thirds of those with low net worth also had low income (15/24), and about one half of those with low income also had low net worth (15/32).

Because the estimates have been adjusted using a scale based on the poverty lines, the dollar amount cutoffs that define the LILNW group differ by household size and, for one- and two-person households, by age of householder. For one-person households with householder age 65 or over, the upper (annualized) income bound for the LILNW group is $5,047, while the upper net worth bound is $9,875. For two-person households with householder age 65 or over, the income bound is $6,369, while the net worth bound is $12,462. The bounds are about nine percent higher if the householder is under age 65. 10/

We are interested here in who the aged LILNW households are and what resources they have. Table 4 shows the composition of the LILNW group compared to all households for the 65 and over and 75 and over age groups. 11/ Looking at sex of householder, 46 percent of all households 65 and over have a female householder, but 75 percent of the LILNW

households have a female householder. For the 75 and over group, 86 percent of the LILNW households have a female householder. For household size, the LILNW group shows substantially more one-person households than the population as a whole in the age group (76 percent compared to 44 percent in the 65 and over group). The LILNW group shows substantially fewer married spouse present householders (11 percent compared to 43 percent in the 65 and over group) and more widowed, divorced, and other householders. Looking at these three characteristics together, households of size one with a widowed female householder comprise 49 percent of the LILNW 65 and over group, but only 27 percent of the total in that age group. The corresponding percentages for the 75 and over group are 68 percent in the LILNW group and only 37 percent overall. Thus, LILNW households tend to be smaller, have female householders more often, and have householders who are widowed, divorced, or other marital status more often than all households in the age group. 12/

Table 5 shows the composition of total income, mean amounts of income types, and the percentage of households with various types of income for all households in the 65 and over and 75 and over groups and for LILNW households in those age groups. Looking at all households 65 and over, OASDI accounts for about one third of total income, while earnings and property income each constitute about one fourth, and pension income accounts for about one tenth. The composition of income differs markedly between LILNW and all households. Looking at the 65 and over group, OASDI benefits are far more important in the LILNW group (about two thirds of total income), although the mean amount of benefits in that group is far

below the mean for all households. SSI is also more important in the LILNW group (12 percent compared to 1 percent). All other income sources are less important in the LILNW group, with earnings (9 percent compared to 25 percent) and property income (2 percent compared to 23 percent) showing large differences. The patterns are similar for the 75 and over group. It would be expected that property income would be a minor income source for the LILNW group since the presence of substantial property income suggests amounts of assets that would move the household out of the LILNW group. Mean total income for the LILNW group is only 29 percent of the mean for all households in the 65 and over group and 32 percent in the 75 and over group. Because median income is quite close to mean income in the LILNW group, but not for all households, the ratio of the LILNW median to the overall median is higher--42 percent for the 65 and over group and 48 percent for the 75 and over group. For both mean and median, there is less difference between the 65 and over and 75 and over groups for the LILNW group than for all households. Panel C of Table 5 shows the percentage of the group that received each type of income. Households in the LILNW group receive SSI much more often (32 percent compared to 11 percent for the 65 and over group), and earnings, property income, and pensions much less often than all households in the age group. The percentages receiving OASDI (93-94 percent for the 65 and over group) show little difference.

Table 6 shows the composition of net worth for LILNW households and all households in the 65 and over and 75 and over age groups. For all households 65 and over, home equity and financial assets each account for

about one third of net worth, while durable goods constitute about one tenth. In contrast, in the LILNW group 65 and over, home equity is about one fifth of net worth and financial assets are about one third, but durable goods constitute almost three fifths. The LILNW mean of each asset is far below the mean for all households in the age group. The LILNW group shows larger shares for liquid financial assets and durable goods, but far lower mean amounts of these assets. The LILNW group shows negligible shares for nonliquid financial assets, business equity, and other assets, and a lower share (and far lower mean) for home equity. Mean and median amounts of net worth are very low in the LILNW group, even below the income amounts. If durable goods are excluded from the mean (on the grounds that that type of asset cannot easily be liquidated or borrowed against), then the mean amount is $1,260 (including home equity).

Panel C of Table 6 shows the percentage of households with each type of asset or debt. Home equity, nonliquid financial assets, business equity, and other assets are held by few households in the LILNW group. Durable goods and liquid financial assets are held by high percentages. Fewer households in the LILNW group have unsecured debt. Differences between the 65 and over and 75 and over groups are relatively small.

V. Summary and Conclusions

In this paper, the income and net worth of aged households in 1979 have been examined using data from the Income Survey Development Program file. A particular emphasis was on the income and net worth of aged households with both relatively low income and relatively low net worth.

Aged households (householder age 65 or over) have relatively low incomes; their mean and median amounts are below the amounts for any other age group. Within that aged group, mean and median incomes fall as age rises. For net worth, however, the mean and median for aged households are above the mean and median for all age groups below 45, and about the same as the 45-54 group. As in the case of income, within the aged group mean and median net worth fall as age rises. When an adjustment is made for household size (using the equivalence scale implicit in the U.S. poverty lines), relative mean income and net worth of aged households are about 25 percent higher than in the unadjusted estimates. There is a substantial amount of dispersion in amounts of both income and net worth among aged households.

The characteristics of all aged households and of aged households with both relatively low income (bottom quintile) and relatively low net worth (bottom two quintiles) were examined. This lower group, which constitutes 15 percent of aged households, is called the "LILNW" group. Compared to all aged households, the LILNW group has higher percentages of female householders, one-person households, and widowed householders. About half of the LILNW group consists of widowed female householders living alone.

The composition of income differs substantially between all aged households and the LILNW group. While social security accounts for about one third of the income of all aged households, it is about two thirds of the income of the LILNW group. SSI is an insignificant proportion for all aged households, but is more than one tenth of the income of the LILNW

group; one third of the LILNW group receives SSI. Earnings are far less important in the LILNW group than for all aged households. The mean income of LILNW households is $3,750, compared to $13,060 for all aged households.

The composition of net worth also differs substantially between LILNW and all aged households. For the LILNW group, almost 60 percent of net worth consists of durable goods; for all aged households only about 10 percent is in that form. Home equity and nonliquid assets are far more important for all aged households than for LILNW households; only a small percentage of LILNW households holds those assets. Mean net worth for the LILNW group is only $3,000 ($1,260 if durable goods are excluded), compared to $79,380 for all aged households.

We have seen that, on the average, aged households have relatively low incomes and relatively high net worth. However, because there is substantial dispersion in these amounts, the averages do not tell the whole story. For example, in 1979 about 2 1/2 million aged households had both relatively low income (mostly OASDI and SSI) and relatively low net worth (mostly durable goods). However, there also are aged households with substantial amounts of both income and net worth. Thus, general statements about how well off "the aged" are can frequently be misleading and should be interpreted with caution.

Table 1.--Mean and Median Income and Net Worth, by Age of Householder, 1979

Age of Householder	Sample Cases	Weighted Number of Households (Thousands)	Annualized Income ($)		Net Worth ($)	
			Mean	Median	Mean	Median
All Ages	6,922	82,211	21,240	16,500	62,430	25,770
Under 25	621	6,613	14,140	12,860	8,880	3,450
25-34	1,441	19,272	20,110	18,000	24,520	12,110
35-44	1,089	14,014	25,380	20,430	64,950	29,820
45-54	1,049	12,975	28,380	24,180	79,120	39,180
55-64	1,155	12,722	25,490	19,330	105,740	51,440
65 and over	1,567	16,614	13,060	8,630	79,380	38,720
65-69	489	5,570	15,560	9,750	90,040	47,300
70-74	411	4,818	12,790	9,040	85,480	45,440
75 and over	667	6,226	11,050	7,460	65,130	28,410

NOTE: All estimates are preliminary. Dollar amounts are rounded to the nearest $10.

Table 2.--Relative Mean Income and Net Worth, by Age of Householder, 1979

	Unadjusted for Size of Household		Adjusted for Size of Household	
Age of Householder	Relative Mean Income	Relative Mean Net Worth	Relative Mean Income	Relative Mean Net Worth
All Ages	1.00	1.00	1.00	1.00
Under 25	.67	.14	.75	.15
25-34	.95	.39	.94	.35
35-44	1.19	1.04	1.01	.83
45-54	1.34	1.27	1.22	1.09
55-64	1.20	1.69	1.27	1.75
65 and over	.62	1.27	.79	1.59
65-69	.73	1.44	.91	1.76
70-74	.60	1.37	.77	1.70
75 and over	.52	1.04	.69	1.35

NOTE: All estimates are preliminary.

Table 3.--Percentage Distribution of Households Among Income and Net Worth Quintiles, Adjusted for Household Size, 1979

Quintiles	All Ages	Age of Householder: 65 and over	Age of Householder: 75 and over
Income			
1	20	32	39
2	20	27	29
3	20	19	14
4	20	12	12
5	20	11	6
Total	100	100	100
Net Worth			
1	20	10	11
2	20	14	17
3	20	16	20
4	20	26	27
5	20	34	25
Total	100	100	100

NOTE: All estimates are preliminary.

Table 4.--Percentage Composition of Aged Households, 1979

	Age of Householder			
	65 and over		75 and over	
Type of Household	Total	Low Income and Low Net Worth	Total	Low Income and Low Net Worth
Total	100	100	100	100
Sex of Householder				
Female	46	75	54	86
Male	54	25	46	14
Household Size				
1 Person	44	76	50	88
2 Persons or more	56	24	50	12
Marital Status of Householder				
Married, Spouse Present	43	11	36	7
Widowed	41	61	55	75
Divorced	6	13	2	4
Other	9	15	9	14
Household Size 1, Householder Widowed Female	27	49	37	68
Sample Cases	1,567	402	667	201
Weighted Number of Households (thousands)	16,614	2,562	6,226	1,213

NOTE: All estimates are preliminary.

Table 5.--Percentage Composition of Income, Mean Incomes, and Percent Receiving Each Income Type, 1979

	Age of Householder			
	65 and over		75 and over	
	Total	Low Income and Low Net Worth	Total	Low Income and Low Net Worth
	A. PERCENTAGE COMPOSITION			
Total	100	100	100	100
Earnings	25	9	17	3
Property	23	2	26	2
OASDI	34	68	39	75
SSI	1	12	2	10
Pension	11	4	9	3
Government	6	4	5	2
Private	5	1	4	1
Other	6	5	7	6
	B. MEAN AMOUNTS ($)			
Total	13,060	3,750	11,050	3,490
Earnings	3,240	330	1,890	120
Property	3,020	60	2,900	60
OASDI	4,400	2,540	4,310	2,620
SSI	190	450	230	360
Pension	1,450	170	980	120
Government	770	130	540	80
Private	680	40	440	40
Other	760	200	740	220
Addendum:				
Median Total Income ($)	8,630	3,630	7,460	3,580
	C. PERCENT RECEIVING EACH INCOME TYPE			
Total	100	100	100	100
Earnings	28	15	21	11
Property	77	36	80	45
OASDI	94	93	97	95
SSI	11	32	13	28
Pension	39	12	33	14
Government	15	6	13	6
Private	28	6	22	8
Other	19	27	22	24

NOTE: All estimates are preliminary. Dollar amounts are rounded to the nearest $10.

Table 6.--Percentage Composition of Net Worth, Mean Amounts of Assets, and Percent Holding Each Type of Asset, 1979

	Age of Householder			
	65 and over		75 and over	
	Total	Low Income and Low Net Worth	Total	Low Income and Low Net Worth
	A. PERCENTAGE COMPOSITION			
Net Worth	100	100	100	100
Wealth	101	108	101	111
Home Equity	32	18	30	9
Financial Assets	35	31	44	43
Liquid	10	31	12	42
Nonliquid	25	1	32	1
Business Equity	7	0	5	0
Other Assets	18	0	13	0
Durable Goods	9	58	8	59
Unsecured Debt	1	8	1	11
	B. MEAN AMOUNTS ($)			
Net Worth	79,380	3,000	65,130	3,110
Wealth	79,930	3,240	65,530	3,460
Home Equity	25,110	550	19,480	290
Financial Assets	28,020	940	28,610	1,320
Liquid	8,020	920	7,560	1,290
Nonliquid	20,000	20	21,040	30
Business Equity	5,660	0	3,400	0
Other Assets	14,090	0	8,640	0
Durable Goods	7,060	1,740	5,410	1,840
Unsecured Debt	540	240	400	350
Addendum:				
Median Net Worth ($)	38,720	2,170	28,410	2,300
	C. PERCENT HOLDING EACH ASSET			
Net Worth	100	97	100	99
Wealth	99	97	100	99
Home Equity	66	8	59	4
Financial Assets	95	82	96	86
Liquid	95	82	96	86
Nonliquid	36	2	31	3
Business Equity	4	0	3	0
Other Assets	18	0	16	0
Durable Goods	98	96	99	99
Unsecured Debt	39	26	28	21

NOTE: All estimates are preliminary. Dollar amounts are rounded to the nearest $10.

FOOTNOTES

*The author is greatly indebted to Sharon Johnson, who prepared the estimates, and to Benjamin Bridges and Denton Vaughan for their helpful comments.

1/ See U.S. Bureau of the Census (1981a) for the CPS definitions.

2/ Net income fom nonfarm self-employment was not obtained in Wave 5 of the ISDP. The "draw" is the amount of salary or money taken out of the business for living expenses.

3/ Those two items were collected in Wave 2 of the ISDP for most households.

4/ The estimates in this paper exclude a few observations with negative household income.

5/ The choice of this scale was arbitrary. However, this is a familiar scale that is much less extreme than a per capita adjustment. The extensive literature on equivalence scales contains many estimated scales (e.g., van der Gaag and Smolensky 1982) that could have been used.

6/ The scale values used are: 1 person (under 65), 1.024; 1 person (65+), 0.943; 2 persons (under 65), 1.322; 2 persons (65+), 1.190; 3 persons, 1.568; 4 persons, 2.009; 5 persons, 2.379; 6 persons, 2.687; 7 persons or more, 3.329.

7/ In these estimates, each household is counted once (i.e., at its sample weight). One alternative would have been to count each household once for each person in the household; that weighting produces relative means that are only slightly different from those shown. See Danziger and Taussig (1979) for a discussion of these weighting issues.

8/ It should be noted that these definitions are somewhat arbitrary. The bottom two quintiles were used instead of the bottom quintile for net worth because the implied net worth amounts in the bottom quintile were extremely small. Looking at these low income and low net worth households does not require a completely specified ranking of households according to economic well-being. Also, we do not claim that all low income and low net worth households are worse off than all other households.

9/ Using estimates unadjusted for household size, 20 percent of aged households were in the LILNW group, compared to 13 percent for all ages (Radner and Vaughan 1984, Table 16). For comparison, in 1979, about 12 percent of all households and 18 percent of aged households were poor (U.S. Bureau of the Census 1982, Table 22).

10/ The bounds for other household sizes can be derived using the scale values shown in footnote 6. The income bounds are about 45 percent above the 1979 poverty lines.

11/ Because of space limitations, estimates for non-LILNW households are not shown separately in this paper. In general, those estimates are similar to the estimates shown for all households in the age group.

12/ As a sensitivity test, the composition of the bottom 16 percent of aged households ranked by the sum of income and net worth after adjustment for household size was examined. Those percentages were quite similar to those shown in Table 4; the alternative ranking showed slightly fewer one-person and widowed households in the lower group. The proportion of the lower group accounted for by widowed females living alone was only about one third, compared to about one half in the LILNW group.

REFERENCES

Danziger, Sheldon and Michael K. Taussig (1979), "The Income Unit and the Anatomy of Income Distribution," Review of Income and Wealth, Series 25, No. 4, (December), pp. 365-375.

Projector, Dorothy S. and Gertrude S. Weiss (1966), Survey of Financial Characteristics of Consumers, Board of Governors of the Federal Reserve System.

Radner, Daniel B. and Denton R. Vaughan (1984), "The Joint Distribution of Wealth and Income for Age Groups, 1979," ORS Working Paper No. 33, Office of Research, Statistics, and International Policy, Social Security Administration (March).

U.S. Bureau of the Census (1981a), Money Income of Families and Persons in the United States: 1979, Current Population Reports, Series P-60, No. 129, Washington.

U.S. Bureau of the Census (1981b), Characteristics of the Population Below the Poverty Level: 1979, Current Population Reports, Series P-60, No. 130, Washington.

U.S. Bureau of the Census (1982), Characteristics of Households and Persons Receiving Selected Noncash Benefits: 1980 (With Comparable Data for 1979), Current Population Reports, Series P-60, No. 131, Washington.

van der Gaag, Jacques and Eugene Smolensky (1982), "True Household Equivalence Scales and Characteristics of the Poor in the United States," Review of Income and Wealth, Series 28, No. 1 (March) pp.17-28.

Weisbrod, Burton A. and Lee W. Hansen (1968), "An Income-Net Worth Approach to Measuring Economic Welfare," American Economic Review, Vol. LVIII, No. 5 (December) pp. 1315-29.

THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

THE SURVEY OF INCOME
AND
PROGRAM PARTICIPATION

No. 8405

Roger A. Herriot &
Daniel Kasprzyk
Bureau of the Census

U.S. Department of Commerce BUREAU OF THE CENSUS

U.S. BUREAU OF THE CENSUS
John G. Keane, Director
C.L. Kincannon, Deputy Director
William P. Butz, Associate Director
for Demographic Fields

POPULATION DIVISION
Roger A. Herriot, Chief

Acknowledgements

This paper was prepared by Roger A. Herriot and Daniel Kasprzyk, Population Division. In this paper, we reviewed work done by Census Bureau staff in many divisions. We appreciate and have benefited from the information provided by our colleagues. Special thanks go to Susan Miskura and John Paletta, Statistical Methods Division, for their review of an early draft. Clerical and editorial assistance was provided by Hazel Beaton, Mary Kisner, and Delma Frankel.

Suggested Citation

Herriot, Roger A., and Kasprzyk, Daniel. "The Survey of Income and Program Participation," SIPP Working Paper Series No. 8405. U.S. Bureau of the Census, Washington, D.C. 1984.

Preface

This paper was featured in the "Case Studies in Panel Survey Design: The International Experience" session, one of three sessions in the Social Statistics Section of the annual meeting of the American Statistical Association.

SIPP is a new Census Bureau survey collecting data that will help measure income distribution and poverty throughout the country more accurately. These data will be used to study Federal and state aid programs (such as food stamps, welfare, Medicaid, and subsidized housing), to estimate future program costs and coverage, and to assess the effects of proposed changes in program eligibility rules or benefit levels.

Households in the survey will be interviewed at 4-month intervals over a period of 2 1/2 years. The reference period will be the 4 months preceding the interview. In all, about 20,000 households will be interviewed, approximately 5,000 each month. Field operations will be handled through our 12 regional offices.

Recurring questions will deal with employment, types of income, and noncash benefits. Periodic questions will be added dealing with school enrollment, marital history, migration, disability, and other topics. Special supplemental questions will also be added to the SIPP questionnaire.

This paper presents the rationale for the survey and the history of its development. It discusses the survey's content, design, frequency, data products, and current research activities.

TABLE OF CONTENTS

	Page
Introduction	1
The Need for a New Survey	1
SIPP Design Features	5
SIPP Collection Procedures	7
SIPP Content: Control Card	9
SIPP Content: Core Data	10
SIPP Content: Fixed Topical Modules	11
SIPP Content: Variable Topical Modules	13
SIPP Data Products	14
SIPP Unit Nonresponse Rates	17
SIPP Related Research: Discussion of SIPP Research Issues	18
SIPP Research: Accessing SIPP Data	19
SIPP Research: Longitudinal Concepts-Household and Family Definition	21
SIPP Research: Longitudinal Estimation	22
SIPP Research: Longitudinal Imputation	23
SIPP Research: Use of Combined Survey and Administrative Data	24
SIPP Research: Panel Surveys as a Source of Migration Data	26
SIPP Research: Wave-to-Wave Changes in Income and Program Receipt	26
SIPP Research: Longitudinal Feedback and Reconciliation System	27
SIPP Research: Sampling for Special Subpopulations	27
SIPP Research: Sampling Error Estimation	28
SIPP Research: The American Statistical Association-Census Bureau Research Fellow Program	28
Summary	29

Figure 1. SIPP Interview Schedule

Figure 2. SIPP Interview Months and Reference Periods

Figure 3. SIPP Income and Asset List

Figure 4. Survey of Income and Program Participation Topical Module Schedule

Figure 5. SIPP Noninterview Data: Wave 1 and 2

Bibliography

The Survey of Income and Program Participation

Introduction

In October 1983, the Bureau of the Census conducted the first interviews of the Survey of Income and Program Participation (SIPP). The SIPP is a nationally representative household survey intended to provide detailed information on all sources of cash and noncash income, eligibility and participation in various government transfer programs, disability, labor force status, assets and liabilities, pension coverage, taxes, and many other items. Data from the survey will provide a multiyear perspective on changes in income, and their relationship to participation in government programs, changes in household composition, and so forth. In general, the SIPP data system is designed to directly measure elements of the tax and transfer system in a comprehensive data base. The purpose of this paper is to review the need for a new survey, briefly describe the research and development work leading up to the SIPP, the survey design features, procedures, and content of the survey. Data products and current survey and research activities will also be described.

The Need for a New Survey

The development of SIPP arose in response to a recognition that the principal source of information on the distribution of household and personal income in the United States, the March Income Supplement of the Current Population Survey (CPS), had severe limitations which could only be rectified by making substantial changes in the survey instrument and procedures. In particular, the CPS collects employment/unemployment information each month and personal, family, and household income once a year. The collection of income information has been and always will be of secondary importance to the goals of this survey. As a consequence, analysts have identified several flaws in the CPS income data.

For example, the CPS:

1) measures annual income but does not collect tax filing status, liabilities, and deductions;
2) does not measure monthly income flows and month-to-month changes;
3) provides annual income estimates, whereas eligibility for most Federal programs is based on a monthly accounting period;
4) produces estimates of last year's income based on current household membership, ignoring moves and changes in household composition during the year;
5) does not measure asset holdings and liabilities and does not provide enough measures of categorical information to produce sound estimates of program eligibility; and
6) underestimates income from transfer programs, retirement and disability income, unemployment compensation, and property income.

These deficiencies are especially serious because the scope of policy analyses was broadened during the 1960's and early 1970's when many public assistance programs were expanded and reorganized. Analysts became more sophisticated in the framing of program policy as well as in the analysis of alterations to programs. In the latter case, microsimulation models were designed to measure the distributional impacts of various welfare reform proposals. Because of the availability of public-use microdata files from large household surveys and high speed computers, policy researchers redirected their questions toward determining the effects of changes in eligibility rules or benefit levels on the different demographic groups. The vigorous exploitation of microsimulation methodology made obvious the data limitations of the CPS. The lack of monthly income data, monthly household composition, and asset information forced model-builders to make many assumptions and imputations of intrayear data in order

to carry out their activities. The assumptions, when invalid, decrease the reliability of estimates of both the number of persons eligible for a program and population subgroups affected by altering program rules and tax policy. In this environment, with analysts needing more detailed data as well as improved measures of cash and noncash income, the Income Survey Development Program (ISDP) began. The purpose of the ISDP, authorized in 1975, was to design and prepare for a major new survey, the Survey of Income and Program Participation (SIPP). The development effort was directed by the Office of the Assistant Secretary for Planning and Evaluation in the Department of Health and Human Services and was carried out jointly with the Bureau of the Census, which assisted in the planning and carried out the field work, and the Social Security Administration (SSA), which administers the major cash income security programs.

The ISDP developed methods intended to overcome the three principal shortcomings of the CPS: 1) the underreporting of property income and other irregular sources of income; 2) the underreporting and misclassification of participation in major income security programs and other types of information that people generally find difficult to report accurately (for example, monthly detail on income earned during the year); and 3) the lack of information necessary to analyze program participation and eligibility.

The principal method by which the CPS shortcomings and the new data requirements were to be addressed was the use of a longitudinal survey design. Persons at sample addresses were interviewed about their income and other characteristics for the previous 3 months. They were then recontacted at regular intervals, having been followed to new addresses if necessary, and asked additional questions to cover the intervening period. Any other persons that they had moved in with, or vice versa, were also interviewed. This continued for 15 months, and ended

with a set of questions on taxes. In this way a highly detailed record was built up for each person and household for an entire calendar year. This design minimized the need for sample persons to recall the details of income and other characteristics for more than a few months and reduced the number of questions that had to be answered in each interview.

Because less time was required to update the basic information after the first interview, time was available in later interview waves to ask questions about other topics that were either stable enough not to require periodic updating--marital history and pension coverage, for example--or emerging issues of one-time interest, such as emergency energy assistance. This design enabled a set of core questions on income and other eligibility determinants to be developed well in advance, thereby, ensuring timely processing and rapid turnaround while leaving interview time to add questions on new policy issues on short notice.

Much of the work of the ISDP centered around four experimental field tests that were conducted to examine different concepts, procedures, questionnaires, recall periods, and the like. 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 the feasibility of many new ideas and methods. 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 8,200 households and provided a vehicle for feasibility tests and controlled experiments of alternative design features. Although used primarily for methodological purposes, it was sufficiently large to provide reliable national estimates of many characteristics of interest

to analysts. (Public-use microdata files and documentation of the 1979 Research Panel are available through the National Technical Information Service. 1/) A more detailed discussion of the ISDP and its activities are provided in Ycas and Lininger (1981) and David (1983).

The ISDP showed that improved income estimation was possible using an ISDP-type survey design, people will participate in a survey with repeated interviews on a subannual basis, the difficulty of collecting income data can be overcome through proper training and motivation, and a national longitudinal survey collecting income and program participation data on a subannual basis is operationally feasible. As a result, it is not surprising that many characteristics of the ISDP are reflected in the SIPP design, including the survey design, content, and questionnaire format.

SIPP Design Features

SIPP started in October 1983 as an ongoing survey program of the Bureau of the Census with one sample panel of approximately 26,000 "designated" households in 174 primary sample units (PSU's) selected to represent the noninstitutional population of the United States. The actual sample size was somewhat smaller (about 21,000 households) because some of the selected households were unoccupied, demolished, converted for nonresidential use, or occupied by persons not eligible for interview, such as persons maintaining a usual residence elsewhere. The sample design is self-weighting; that is, each unit selected in the sample has the same probability of selection.

Each household is interviewed once every 4 months for 2 1/2 years to produce sufficient data for longitudinal analyses while providing a relatively short

---

1/ To request tapes and documentation describing the history of the 1979 Research Panel, sample design, survey content, estimation procedures, and data collection and processing procedures, write to: Department of Commerce, National Technical Information Service, 5385 Port Royal Road, Springfield, Virginia 22161 or call (703) 487-4807.

recall period for reporting monthly income. The reference period for the principal survey items is the 4 months preceding the interview. For example, in October, the reference period is June through September; when the household is interviewed again in February, it is October through January. This interviewing plan will result in eight interviews per household.

In January 1985 and every January thereafter, a new, slightly smaller panel will be introduced. (Figure 1 illustrates the plan for implementing sample panels in 1984 through 1987.) This design will allow cross-sectional estimates to be produced from a combined sample of approximately 35,000 households. The overlapping panel design will also enhance the estimates of change, particularly year-to-year change. 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 group of households. This design will give a more precise estimate of change than a similar comparison involving two different groups of households in which greater sampling variability obscures the actual change.

Finally, to facilitate field operations, each sample panel is divided into four approximately equal subsamples, called rotation groups; one rotation group will be interviewed in a given month. Thus, one cycle or "wave" of interviewing takes 4 consecutive months. This design creates manageable interviewing and processing workloads each month instead of one large workload every 4 months; however, it results in each rotation group using a different reference period. Figure 2 provides an illustration of the relationship between waves, rotation groups, interview months, and reference periods after the sample is fully operational in 1985.

To recap, the panels are the most important feature of the SIPP design--a new panel is initiated each year in January with a sample size of about 20,000 households. There are also waves or interviews in each panel. Each panel, other than the 1984 which is idiosyncratic, consists of eight waves of interviews. Finally, there are rotation groups within each wave. Four rotation groups comprise a wave, each consisting of approximately one-fourth of the total sample, that is interviewed each month. Since the reference period is the 4 months prior to the interview, one should notice that each rotation group has a different reference period.

The collection of data on a "staggered" basis produces 7 months of data under a 4-month reference period. This occurs because a full sample of cases during the wave is not available for each month of the reference period. Looking again at figure 2, January interviews collect data for the period September through December; February interviews, for October through January; March interviews, for November through February; and April interviews, for December through March. By considering the use of individual wave files, the September data will only be available for the first rotation group; the October data, the first two rotation groups; the November data, the first three rotation groups; and so forth. Because of the design, however, matching individual wave data together will allow monthly analysis on the full sample. Although SIPP is essentially a monthly survey, the staggered design and the consequently staggered reference period only allow analysis on a full sample for one month of each wave.

SIPP Collection Procedures

Data collection operations are managed through the Census Bureau's 12 permanent regional offices. Interviewers assigned to these offices conduct one personal visit interview with each sampled household every 4 months. At the time of

the interviewer's visit, each person 15 years old or older who is present is asked to provide information about himself/herself; a proxy respondent is asked to provide information for those who are not available. Telephone interviewing is permitted only to obtain missing information or to interview persons who will not or cannot participate otherwise.

The average length of the interview is about 30 minutes. An important design feature of SIPP is that all persons in a sampled household at the time of the first interview remain in the sample even if they move to a new address during the next 2 1/2 years. For cost and operational reasons, person-visit interviews are only conducted at new addresses that are in or within 100 miles of a SIPP primary sampling unit. After the first interview, the SIPP sample is a person-based sample, consisting of all individuals who were living in the sample unit at the time of the Wave 1 interview. Individuals aged 15 and over who subsequently share living quarters with the original sample people will also be interviewed in order to provide the overall economic context of the original sample persons. Changes in household composition caused by persons who join or leave the household after the first interview are also recorded. These individuals are interviewed as long as they reside with an original sample person.

Another important feature of SIPP is the identification numbering system. Each person will be assigned a unique fourteen-digit identification (ID) number at the time he/she enters the sample; an additional two-digits will be assigned if the person moves to a new address. A master list of identification numbers will be used by the regional offices to monitor the status of interviewing each month after Wave 1. The regional offices will be responsible for ensuring that there is a completed questionnaire (or reason for noninterview) for each number on the list representing all the persons assigned for interview in a month.

The list will be updated regularly to account for persons who are added or deleted from the sample.

The ID is intended to provide a means of linking information about an individual across time and uniquely identifying which household each person is a member of at any point in the panel. Through the ID system, we expect to link data from all persons ever associated with a given household throughout the 2 1/2-year duration of a panel. This will facilitate the construction of household income estimates based on the actual composition of households during the measurement period. More information about the construction and use of the ID number can be found in Nelson, McMillen, and Kasprzyk (1984) and Jean and McArthur (1984).

SIPP Content: Control Card

The control card is used to obtain and maintain information on the basic characteristics associated with households and persons and to record information for operational control purposes. Characteristics recorded on the control card by the interviewer include the age, race, ethnic origin, sex, marital status, and educational level of each member of the household, as well as information on the housing unit and the relationship of the householder to other members. A household respondent provides this information, which is updated as appropriate, at each interview. The control card is also used to keep track of when and why persons enter and leave the household, thereby providing enough information to automatically create monthly household and family groups. There is also space to record information that will improve our ability to follow persons who move during the survey. In addition, after each visit, data on employment, income, and other information is transcribed from the core questionnaire to the control card so the data can be used in the next interview. Finally, social security numbers for every person in the sample unit will be recorded on the SIPP control card.

SIPP Content: Core Data

The content of SIPP was developed around a "core" of labor force and income questions designed to measure the economic situation of persons in the United States. These questions expand the data currently available on the distribution of cash and noncash income and are repeated at each interviewing wave. SIPP core data build an income profile of each person aged 15 and over in a sample household. The profile is developed by determining the labor force participation status of each person in the sample and asking specific questions about the types of income received, including transfer payments and noncash benefits from various programs for each month of the reference period. A few questions on private health insurance coverage are also included in the core.

Persons employed at anytime during the 4-month reference period are asked to report on jobs held or businesses owned, number of hours and weeks worked, hourly rate of pay, amount of earnings received, and weeks without a job or business. Questions for those who were not working some or all of the 4-month reference period are designed to determine if they should be classified as unemployed, discouraged workers, or not in the labor force.

In addition to questions about labor force activity and the earnings from a job, self-employment, or farm, the core includes questions related to nearly 50 other types of income. Questions about common income types are specifically asked while the receipt of less common income types are elicited through general probing questions. Questions are asked about the receipt of government transfer payments from retirement, disability, unemployment benefits, and welfare programs such as food stamps and Aid to Families with Dependent Children (AFDC). Information on the receipt of noncash benefits from programs such as Medicare and Medicaid is also obtained. Other income questions in the

core relate to private transfers such as pensions from employers, alimony, and child support. For certain income types, such as food stamps and AFDC, questions are included which help to identify the household members covered by the payment, thus allowing the proper construction of program analysis units. Finally, the core data also includes questions on the ownership of assets which produce income, such as savings accounts, money market accounts, NOW accounts, stocks, mutual fund shares, and rental property. The amounts of income received from these income producing assets are also obtained, as well as indications of joint holdings and estimates of account balances if the amount of interest is not known. A list of income and asset types collected in the SIPP is provided in figure 3.

SIPP Content: Fixed Topical Modules

The core data provide information on a continuing basis about levels of economic well-being and changes in these levels over time. These data, while extremely detailed, allow analyses of well-being which only account for income and demographic variables. The SIPP has been designed to provide a broader context for analysis by adding questions on a variety of topics not covered in the core section. These questions are labelled "fixed topical modules" and are assigned to particular interviewing waves of the survey. If more than one observation is needed, questions on one wave may be repeated on a later wave.

The administration of these modules of questions is made possible by the fact that less time is required to update the core information collected in the first interview. Also, the topics covered in these modules do not require repeated measurement at each interview and, therefore, may use a reference period longer than the period used for the information obtained in the core. For example, the third SIPP interview questionnaire collects information on health and

disability, and education and work history. The former are obtained because they are among the major factors affecting a person's ability to work, his/her earnings, sources of income, and participation in public programs. The latter provide data to help understand a person's economic situation in relationship to his educational and occupational background. The fourth interview contains topical data on assets and liabilities, retirement and pension coverage, and housing conditions/energy usage. The collection of assets and liabilities data allows the study of economic well-being beyond that which can be observed through the study of income alone. Participation in many Federal programs is contingent upon assets held by the individual or household, as well as income. Some of the major types of assets collected are dividend earnings, home equity, durable goods, and unsecured liabilities. This module, unlike the core data on assets, collects the value of the assets such as the market value of real estate, vehicles owned, and stocks held. This module will be administered twice in each panel, in waves one year apart.

A module of questions on tax-related information will be fielded in the sixth interview. Information on filing status and taxes paid (income, property, and social security (FICA) taxes) will be obtained to allow the estimation of tax incidence, disposable income, and the simulation of tax policy alternatives. In addition to tax-related matters, an annual "round-up" module will be administered in the fifth and eighth interviews (sixth and ninth interviews in the 1984 Panel) to obtain wage and salary data from W-2 forms and estimates of annual self-employment and property income for each appropriate person.

Other fixed topical modules which will be administered in only one wave of the survey include: 1) educational enrollment; and 2) marital history, fertility, and migration. A schedule for implementation of the topical modules is given in figure 4.

SIPP Content: Variable Topical Modules

In response to program planning and policy analysis data requirements, the final component of the SIPP content consists of modules of questions designed in consultation with other Federal agencies. These modules will not be a recurring feature of each SIPP panel. Question modules have been identified for the first SIPP panel and the first year of the second panel; their schedule for implementation is given in figure 4.

Brief descriptions of several of these variable topical modules will indicate the breadth of information collected. For example, in the fourth interview, a retirement and pension module will contain questions on coverage and vested rights in retirement and pension plans. The data will help in the analysis of how net worth is related to retirement decisions and will provide a comparison of the Social Security System with private retirement plans.

Another topical module in the fourth interview will collect characteristics of households that affect energy usage. These data will help provide a better estimate of income remaining after all housing needs are met and fulfill a need for information concerning energy usage, as well as provide information to allow the simulation and analysis of individuals and households qualifying for energy assistance programs.

In the fifth interview, a child care topical module has been 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 because child care expenses are a major part of work-related expenses and are frequently deductible for program eligibility purposes.

In the same interview, questions on welfare history and child support will help determine the length of time persons receive aid from specific welfare programs, as well as provide information on child support agreements. The data from the welfare history questions will measure the extent to which persons and households have been dependent upon government transfer programs. Questions concerning child support will measure the degree to which the failure of the father to provide child care affects the likelihood of the mother and children's participation in government transfer programs.

A topical module on reasons for not working/reservation wage will include questions to ascertain why persons are not in the labor force. The data collected will aid the understanding of the conditions required for unemployed persons to accept a job.

Yet another topical module will contain questions about providing regular payments for the support of persons who are not members of the SIPP household and about expenses associated with a person's job. Those questions will help in obtaining a measure of the fixed financial obligations of persons, resulting in a more complete picture of their economic situation.

As can be seen, a wide variety of topics are covered under the aegis of the variable topical module concept. The breadth of these data in combination with the income and asset information and the fact that the data requirements of these modules arose from specific policy requirements ensure that SIPP will be a widely used and powerful data base serving multiple purposes.

SIPP Data Products

A number of publications and public-use data files will be generated from the information collected in SIPP. Both publications and data files are identified

by whether they are cross-sectional or longitudinal. Two types of cross-sectional reports are planned by the Census Bureau: 1) a set of quarterly and annual reports that will focus on core information; and 2) a set of periodic or single-time reports that will use the detailed data from the topical modules.

The quarterly cross-sectional reports will provide average values for a variety of labor force, income, and household composition measures based on monthly averages. The first quarterly report will be issued in fall 1984 and will contain income and labor force data referring to the third quarter of 1983. The annual reports will be similar in content, but will show values averaged across 12 months rather than 3 months. The periodic and single-time reports will use the detailed data from the topical modules to examine issues related to income and program participation. These reports may also focus solely on the material covered in a topical module such as work history or migration.

Plans for longitudinal data reports are under discussion, but they are expected to concentrate on data that can be used to examine trends and changes over time. This 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 reports have been proposed for consideration (McMillen and Kasprzyk (1984)):

1) economic profile reports, presenting yearly aggregates of monthly data on individuals--reporting household and family information as characteristics of 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 spells or experiences in a given state as well as the duration of those spells;
5) longitudinal family and unrelated individual reports, presenting the characteristics of longitudinal family units defined in SIPP (see McMillen and Herriot (1984) for more information on this topic); and
6) special event reports, providing data related to a unique event, such as marriage, divorce, separation, the birth of a child, a return to school, or a move to a new address; these reports could also use multiwave core data combined with topical module data such as wealth or work history.

SIPP cross-sectional data files will be issued on a wave-by-wave basis. Each file will include person, family, and household information collected in the survey wave. Almost all data obtained on the questionnaire will be included on the files; certain summary income recodes will also be included. Data that might disclose the identity of a person will be excluded or recoded in accordance with standard Census Bureau confidentiality restrictions. Wave files will be edited, imputed, and weighted in a manner consistent with their use for cross-sectional analysis. A unique identification number will be included to allow users to merge two or more SIPP files. However, since the processing of wave files is independent, wave-to-wave data inconsistencies will occur and the user must be prepared to resolve them.

Plans for producing public-use files designed for longitudinal analysis are less well-defined at this time. The basic longitudinal file for SIPP is a calendar-year file of core income data; that is, the data items on the file will be monthly observations for 12 months of the calendar year for each person in the sample. Thus, the first of the series would be a calendar-year (CY) 1984 file from the 1984 Panel. The second of the series would be a CY 1985 file produced in two parts--a CY85 file from the 1985 Panel and a CY85 file from the 1984 Panel. These files will be prepared in such a manner that they can be used separately or with a single adjustment to the weights together. The calendar-year files will be person-level files with monthly amounts of income and program-related data (core data) as well as information about the kinds of program-recipient units to which the individual belonged during the period the file covers. These files will also include longitudinal household and family membership using the "shared experience" definition of longitudinal households (McMillen and Herriot (1984)).

Finally, longitudinal files should be developed separately for <u>each</u> panel. Thus, a 1984 Full Panel file should be created containing core data from Waves 1 through 9 as well as the topical material. Similarly, a 1985 Full Panel file should be created containing core and topical data from all waves.

<u>SIPP Unit Nonresponse Rates</u>

The first SIPP interviews were conducted in October 1983. At this time, cross-sectional unit noninterview rates are available for the first two waves of SIPP. Unit noninterview rates (type A rate) provide a measure of the success/failure of the SIPP field work. While refusals are the largest part of the type A rate, it also includes "no one home" and "temporarily absent" households. Figure 5 provides an overall look at these rates. In Wave 1 (all rotation groups), the

mean type A rate was 4.8 percent; in Wave 2, 3.7 percent. These rates are an improvement on the rates experienced in the Income Survey Development Program and are comparable to the rates obtained by the Current Population Survey.

SIPP-Related Research: Discussion of SIPP Research Issues

Interruptions in the funding support for SIPP in 1982 resulted in a cessation of the analysis associated with the pilot surveys of the development program. These surveys provided a large body of data to address a number of important methodological and substantive issues. Many of these issues were raised by Kasprzyk (1983) and discussed further in such forums as the Office of Management and Budget's (OMB) SIPP Advisory Committee, the Social Science Research Council's Subcommittee on the SIPP, the staff of the Committee on National Statistics, and the American Statistical Association's Census Advisory Committee. This year's meetings of the American Statistical Association are being used to bring the research community up-to-date on a variety of SIPP-related research issues. A wide range of topics, both methodological and substantive, are covered in four sessions organized under the auspices of the Social Statistics and Survey Research Methods Sections.

Participation by SIPP staff in various workshops has also served to enhance and further the discussion of SIPP research. For example, Census Bureau staff participated in the Social Science Research Council (SSRC) Workshop on SIPP, held at the conclusion of the National Bureau of Economics Research Conference on Income and Wealth, and will participate in a Census Bureau - SSRC Workshop on SIPP issues and products during the 1984 meetings of the American Sociological Association. In addition, less formal workshops on longitudinal issues as they pertain to SIPP--analysis concepts and estimation--and on the processing of SIPP longitudinal data have been held to exchange and solicit information from interested analysts.

Finally, a SIPP working paper series has been established as a mechanism to provide timely and widespread access to information developed as part of the SIPP. Papers in the series will cover a broad range of topics including:

1) procedural information on the collection and processing of data;
2) survey methodology research; and
3) preliminary substantive results, such as the measurement of household composition change over time.

A substantial effort has been made by the SIPP staff to exchange ideas related to SIPP research with the research community. A continuing dialogue has been encouraged and expected. Several topics which have been discussed will be described below.

SIPP Research: Accessing SIPP Data

Processing experience with data collected during the development program has shown that the complexity of the data, especially its longitudinal aspects, results in severe difficulties for the analysts. Indeed, the structure of the SIPP cross-sectional data files will seem very complicated to most users. The structure, described by Fink (1984), consists of a five-level hierarchy--sample unit, address or household, family, person, income types--with multiple record types in the fifth or lowest level. This structure, while chosen to provide maximum flexibility for cross-sectional data analysis and to simplify the merging of multiple waves of data into a longitudinal data base, does suggest that alternative ways of accessing SIPP data are necessary.

Acknowledging the data access difficulties, Census Bureau staff has been working with an OMB subcommittee comprised of representatives from various Federal agencies. This subcommittee has proposed the development of an

"alternative" SIPP cross-sectional file. The proposed file has a rectangular structure and the individual is the basic unit of analysis. At this time it seems that the content of the rectangular file will not differ substantially from the more complicated file, although certain kinds of analysis involving households and families will be somewhat more difficult to conduct.

An internal Census Bureau committee investigated the need for and use of a data base management system for SIPP within the Census Bureau's current processing environment. The study concluded that Census Bureau options in the statistical data base field were quite limited and that, of the data base management systems (DBMS) available at the Census Bureau, the Scientific Information Retrieval (SIR) DBMS seemed to be the most suitable for SIPP. As a result of this study, SIPP data are now being structured in the SIR data base format by the Census Bureau's Systems Support Division.

Simultaneously, along with the SIR data base construction, consideration is being given to the development of a software system to accompany SIPP data files and allow the relatively easy generation of extract files of records focusing on changes in status, duration of status, and the number of spells of a particular status. These extract files would serve as an intermediate product to be used as input into the more widely available statistical software packages.

The software system, although still under discussion, has four general characteristics. First, it is designed to provide the user easy access to SIPP data; second, it extracts from the SIPP data base both the time-dimensioned data and summary statistics associated with those data, such as spell duration, number of spells, and number of types of transitions; third, the system allows the user to define both the time period to be considered and the recodes to be used; and fourth, it outputs a data set easily handled by most statistical software.

Finally, to improve the understanding of and access to SIPP data, a data product and delivery program is being developed. This program would include introductory SIPP workshops and user workshops, newsletters, and guides to increase the understanding of SIPP data products.

SIPP Research: Longitudinal Concepts-Household and Family Definition

Household and family level analysis in a longitudinal survey is complicated by the fact that the composition of households and families can change over time since original sample persons leave to join other households or families, or to set up new ones. The principal issue is the development of definitions of households and families which account for survey measurements at two or more points in time and which do not create serious conflicts with the traditional cross-sectional household and family constructs. McMillen and Herriot (1984) discuss this problem in detail by examining cross-sectional concepts and their deficiencies, examining and evaluating several longitudinal concepts, proposing a concept for SIPP, and then illustrating how the concept would be used in calculating aggregate household/family characteristics and in tabulations of the number of households, household types, and characteristics.

The authors conclude that a longitudinal definition of family should be developed separately from that of nonfamily households and that the most reasonable and intuitive longitudinal concept is one based on the "shared experience" of the householder or spouse of the householder. Thus, they propose that a longitudinal family be defined as two or more related persons--at least one of whom is the householder or spouse of the householder--who had the same household experience for two or more consecutive months. No more than one core family unit with children can continue from a previous month family. To distinguish situations where both parents and children split into two more households, three levels of

criteria are suggested: 1) the family with the most child-months is identified as continuous; 2) if two families have the same number of child-months, the family with the most family-months is identified as continuous; and 3) if a distinction cannot be made based on the first two criteria, then the continuing unit will be assigned randomly. Having reached this conclusion, discussion will now focus on the concept of longitudinal nonfamily households.

The issue of the longitudinal concept, although difficult in and of itself, is just the first step in the resolution of several other topics, including tabulation and estimation, also discussed by McMillen and Herriot (1984).

SIPP Research: Longitudinal Estimation

Weighting the longitudinal sample, especially for analytic units other than the individual, is a difficult area requiring a continuing statistician-analyst dialogue. Detailed longitudinal weighting procedures will not be developed until closure is reached on definitions for longitudinal analysis units. Despite the lack of closure, research on this topic has proceeded along two dimensions--longitudinal person estimation and longitudinal household (family or recipient unit) estimation. Since, at a minimum, the first SIPP longitudinal microdata products will be files of person-based information with household (family or recipient unit) as an attribute of the person, the early emphasis of the work is on the longitudinal person estimation. The work includes the calculation of selection probabilities to yield unbiased longitudinal estimates of individual characteristics and the use of controls in additional stages of estimation. Preliminary thoughts on this topic are described by Judkins, Hubble, Dorsch, McMillen, and Ernst (1984).

The topic of longitudinal household (family or recipient unit) estimation is also under study. The work includes the development of an explicit statement

of the longitudinal estimation problem, the survey universe under several definitions of longitudinal household, and the principal estimation issues which require study. This includes, as in the case of person estimation, the calculation of selection probabilities to yield unbiased longitudinal estimates and the use of controls in additional stages of estimation. Alternative approaches to obtaining base weights for longitudinal households (families) and methods of ratio-adjusting the unbiased estimates will also be explored. Preliminary thoughts on this topic are discussed in Ernst, Hubble, and Judkins (1984).

SIPP Research: Longitudinal Imputation

The varieties of types of nonresponse--unit nonresponse defined as nonresponse to all waves of the survey, wave nonresponse defined as nonresponse to a particular wave interview, and item nonresponse defined as nonresponse to a particular item--and their patterns pose numerous difficulties for designing appropriate strategies for nonresponse compensation. At this time several aspects of the problem are being developed. The Statistical Research Division at the Census Bureau is studying patterns of item "missingness"; frequency of state-to-state transitions; and evaluating proposed imputation strategies, both the model-based and hot-deck types, for labor force, wages, and salary items on the 1979 ISDP Panel. Other issues to be studied include: 1) the effect of sort order on survey estimates when using a hot-deck or matching approach; 2) the effect of different collapsing strategies for the hot-deck approach; and 3) the identification of important covariates for modelling and matching. Some preliminary ideas on the treatment of multiwave item nonresponse in the SIPP are discussed in Samuhel and Huggins (1984).

In addition, two other aspects of the problem are also under development:

1) To study item nonresponse for property and program income--in

particular, examining levels and patterns of item nonresponse and developing and comparing methods to treat this nonresponse in these subject areas using more than one wave of data. This work is similar to the previously discussed work on labor force, wages, and salaries, but in a different subject-matter area.

2) To study general strategies of handling missing data in panel surveys, including: (a) a discussion and analysis of issues pertaining to weighting versus imputing for attrition cases; (b) a discussion of the treatment (weighting or imputation) of the so-called "non-nested" missing data cases; (c) an empirical examination and comparison of weighting and imputation using data from the 1979 ISDP Panel; and (d) a discussion based on empirical results of when to choose one strategy over the other.

SIPP Research: Use of Combined Survey and Administrative Data

During its development period, SIPP had been viewed as an integrated data system, combining survey data with administrative record data. Because of the emphasis on this approach, an internal Census Bureau committee was formed to assess and make recommendations regarding the potential uses of administrative records, the development of demonstration pilot studies, and the special confidentiality or privacy issues involved in the use of administrative records. The committee developed a proposal, later implemented, to validate electronically reported social security numbers (SSN), to manually search for SSN's not reported correctly, and to use the panel aspect of the SIPP to correct and verify a respondent's SSN. Having established the link for matching activities, work is now proceeding on identifying content and availability of administrative record systems for use in: a) data augmentation for research and estimates; and b) survey data evaluation. In the former, a working group is developing a research

plan and methodology proposal in which SIPP demographic data would be merged with economic data from census files to form a microdata base for both individuals and the firms in which they work. This project includes: a) identification of the study areas and issues to which the data base would be applied; b) determining the availability, coverage, and quality of economic data in various census files; c) specifying the demographic and economic data contained in the new data base; and d) resolving methodological problems encountered in developing a microfile containing worker and firm data. More information concerning this work can be found in Haber, Ryscavage, Sater, and Valdisera (1984).

Another area of research with respect to administrative record systems is the development of criterion measures for validation studies of items common to both the survey and administrative records. The goal of the project is the improved understanding of the quality of the SIPP data and, ultimately, the development of quantitative estimates of response and nonresponse errors for the purposes of adjusting survey data or modifying survey procedures to obtain better quality data. The first general aspect of this work, which will be undertaken during the next several months, is the development of a research plan identifying activities required to evaluate SIPP data using administrative records. Topics covered in the plan include: a) an investigation of the kinds of record systems which should be considered for survey data evaluation studies; b) alternative strategies to obtain administrative records, particularly state-held files; c) variables to be considered for evaluation; and d) access issues as they pertain to merged administrative and survey data. The second general aspect of this work is the development of a demonstration and feasibility study to examine response and nonresponse errors from Waves 1 and 2 of the SIPP using administrative record systems for the major transfer programs from a limited number of states.

SIPP Research: Panel Surveys as a Source of Migration Data

SIPP data can be used to address a wide variety of migration topics in two ways: 1) by traditional--cross-sectional--analyses, these data serve to further our understanding of how geographical mobility leads to adjustments in labor markets, housing markets, etc.; and 2) the survey's longitudinal design provides a natural source of geographical mobility data because individuals are followed should they move to a new place of residence. Detailed information on the situation of individuals and households at both origin and destination locales will markedly enhance our understanding of the process and effects of geographical mobility. The first stage of this work is to review analyses of migration from previous panel surveys and then assess how SIPP can further our understanding of geographical mobility processes. Some results from this work are described by Dahmann (1984).

SIPP Research: Wave-to-Wave Changes in Income and Program Receipt

Analysts with an interest in the quality of the data obtained in the 1979 ISDP Panel--and the subsequent SIPP--have expressed concern that respondent reports of income receipt may be flawed. There appears to be a tendency for reported program turnover to occur between waves more often than within waves--that is, in the pilot surveys of the development program, between months 3 and 4 rather than the other 4 consecutive pairs of months. It is assumed that this probably represents response error arising from imperfect recall although other factors (for example, mismatching in linking the data files) might also account for this effect. Some analyses of this phenomenon using five waves of data from the 1979 Panel and examining the extent to which differences are related to respondent status patterns across survey waves are presented in Moore and Kasprzyk (1984). Additional work using SIPP data is planned as soon as data are available.

SIPP Research: Longitudinal Feedback and Reconciliation System

Because of its design, SIPP has a potential for missing and inconsistent data problems from wave to wave. The issue of concern is the development of appropriate forms and procedures to identify and correct longitudinal data problems during the collection rather than the processing phase. An automated income and work experience profile to identify potential cross-wave edit failures and data problems could help in the development of SIPP longitudinal data products. This profile would contain responses on labor force activity and amounts of income and program benefits received during the previous calendar year. It would be reviewed for accuracy by each respondent at the conclusion of the calendar year. A system such as this should clarify apparent inconsistent responses by using previously reported amounts to identify and reconcile cross-wave inconsistencies.

Prior to designing and implementing a reconciliation system to smooth transition data during a calendar year, a preliminary review of a subsample of questionnaires for the first two waves of SIPP will be conducted.

SIPP Research: Sampling for Special Subpopulations

After the first year of the 1984 SIPP Panel, Federal program agencies should become interested in adding sample cases for specific subpopulations of interest, such as the high and low income groups, Blacks and Hispanics, and the aged and disabled, to policy analysis. A multidivisional work group is discussing methods for oversampling special populations. In particular, a variety of subsampling (screening) proposals will be analyzed. Issues under discussion are both operational and statistical. In the former, the use of different screening mechanisms, such as the questionnaire (edited or unedited), the control card, or data from the master segment tape, will be discussed. In addition, an investigation into the quality of the data available for subsampling will be discussed.

A statistical issue under consideration is the reliability of estimates when different subsampling schemes are introduced. The work group intends to identify--in a broad sense--characteristics of specific subsampling plans and relate these to changes in reliability. Finally, as the group's work continues, costs and operational feasibility will be ascertained as part of the analysis.

During the last several months, subsampling characteristics based on income and demographic variables have been identified, and estimates of reliability have been obtained for different subsampling rates and different characteristics. A summary of results is now in preparation.

SIPP Research: Sampling Error Estimation

Applications of SIPP data are expected for a wide variety of analyses--micro-simulation modeling, multivariate analysis, and simple tabulations and cross-tabulations. Some topics under study in the Statistical Research Division are: 1) an investigation of currently available computer software which provide general procedures for computing sampling error estimates for a variety of complex survey designs; and 2) an assessment of Census Bureau procedures for computing and estimating the median and its variance.

SIPP Research: The American Statistical Association-Census Bureau Research Fellow Program

Recognizing that SIPP research cannot be confined solely within the Census Bureau, SIPP research planning has encouraged the use of development program data to increase the understanding of the SIPP data and the inherent difficulties working with this new data base. One aspect of this activity has been the expansion of the ASA-Census research fellow program to identify explicitly

SIPP-related research activities. As a result, two research fellows have been chosen for the 1984-1985 academic year: Harold Watts (Columbia University) and Constance Citro (National Academy of Sciences and Mathematica Policy Research). Dr. Watts intends to use the 1979 Panel data to understand changes in living arrangements and how long they last. He is interested in short-term gross flow patterns of change in household status and will characterize the changes in the household status of the individual from one wave to another. Dr. Citro intends to simulate alternative definitions of household continuity. She will develop tables of annual income and poverty levels under alternative definitions using the data from the development program's 1979 Panel and will help interpret the meaning and usefulness of the alternative measures.

Summary

This paper has reviewed the development and current activities of the Survey of Income and Program Participation. Survey design features, selected field procedures, content of the survey, and data products have been described. As a number of methodological issues remain unresolved, selected research activities have been described. While covering many aspects of the SIPP, this paper was not meant to be comprehensive. The scope of topics discussed, however, illustrates the nature, breadth, and opportunities of the Survey of Income and Program Participation.

Figure 2. SIPP Interview Months and Reference Periods

Wave	Rotation Group	Interview Month	Reference Periods																		
							1st Qtr Date			2nd Qtr Date			3rd Qtr Date			4th Qtr Date					
1	1	JAN	SEP	OCT	NOV	DEC															
	2	FEB		OCT	NOV	DEC	JAN														
	3	MAR			NOV	DEC	JAN	FEB													
	4	APR				DEC	JAN	FEB	MAR												
2	1	MAY					JAN	FEB	MAR	APR											
	2	JUN						FEB	MAR	APR	MAY										
	3	JUL							MAR	APR	MAY	JUN									
	4	AUG								APR	MAY	JUN	JUL								
3	1	SEP									MAY	JUN	JUL	AUG							
	2	OCT										JUN	JUL	AUG	SEP						
	3	NOV											JUL	AUG	SEP	OCT					
	4	DEC												AUG	SEP	OCT	NOV				
4	1	JAN													SEP	OCT	NOV	DEC			
	2	FEB														OCT	NOV	DEC	JAN		
	3	MAR															NOV	DEC	JAN	FEB	
	4	APR																DEC	JAN	FEB	MAR
	1	MAY																	JAN	FEB	MAR	APR

Figure 3. SIPP Income and Asset List

INCOME LIST Type	ASSET LIST Type
Social Security	Regular/Passbook Savings Accounts in a Bank, Savings and Loan or Credit Union
U.S. Government Railroad Retirement Pay	Money Market Deposit Accounts
Federal Supplemental Security Income (SSI)	Certificates of Deposit or Other Savings Certificates
State Supplemental Security Income (State Administered SSI Only)	NOW, Super NOW, or Other Interest-Earning Checking Accounts
State Unemployment Compensation	Money Market Funds
Supplemental Unemployment Benefits	U.S. Government Securities
Other Unemployment Compensation (Trade Adjustment Act Benefits, Strike Pay, Other)	Municipal or Corporate Bonds
Veterans Compensation or Pension	Other Interest-Earning Assets
Black Lung Payments	Stocks or Mutual Funds Shares
Worker's Compensation	Rental Property
State Temporary Sickness or Disability Benefits	Mortgages
Employer or Union Temporary Sickness Policy	Royalties
Payments from a Sickness, Accident, or Disability Insurance Policy Purchased on Your Own	Other Financial Investments
Aid to Families with Dependent Children (AFDC, ADC)	
General Assistance or General Relief	
Indian, Cuban, or Refugee Assistance	
Foster Child Care Payments	
Other Welfare	
WIC (Women, Infants, and Children Nutrition Program)	
Food Stamps	
Child Support Payments	
Alimony Payments	
Pension from Company or Unión	
Federal Civil Service or Other Federal Civilian Employee Pensions	
U.S. Military Retirement Pay	
National Guard or Reserve Forces Retirement	
State Government Pensions	
Local Government Pensions	
Income from Paid-Up Life Insurance Policies or Annuities	
Estates and Trusts	
Other Payments for Retirement, Disability, or Survivor	
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	

Figure 4. Survey of Income and Program Participation - Topical Module Schedule

INTERVIEW DATES	1984 Panel			1985 Panel		
	Wave	Fixed Topical Module	Variable Topical Module	Wave	Fixed Topical Module	Variable Topical Module
Oct. 83-Jan. 84	1	None	None			
Feb. 84-Apr. 84	2	None	None			
Mar. 84-Aug. 84	3	Health and Disability Work History Education History	Health Care Utilization and Planning Social Serivces in Kind-Health Social Services in Kind-Training			
Sept. 84-Dec. 84	4	Assets Liabilities	Pension and Retirement Issues Housing Conditions and Cost Energy Usage			
Jan. 85-Apr. 85	5		Child Care Arrangements and Planning Duration of Welfare Support for Non-Household Members Social Services in Kind-Child Care & Other Reasons for Not Working-Reservation Wage Work Related Expenses	1	None	None
May 85-Aug. 85	6	Annual Income Taxes Employee Benefits Education Enrollment	Educational Financing Training Questions (ETA)	2	Education Enrollment	Educational Financing Training Questions (ETA)
Sept. 85-Dec. 85	7	Assets Liabilities	Housing Conditions & Cost Neighborhood Conditions Energy Usage	3	Assets Liabilities	Housing Conditions & Cost Neighborhood Conditions Energy Usage
Jan. 86-Apr. 86	8	Marital History Fertility Migration	Support for Non-Household Members Reasons for Not Working-Reservation Wage	4	Marital History Fertility Migration	Support for Non-Household Members Reasons for Not Working-Reservation Wage
May 86 Aug. 86	9	Annual Income Taxes Employee Benefits Educational Enrollment	Educational Financing Health and Disability Work Related Expenses Retirement Plans and Expectations Training Questions (ETA)	5	Annual Income Taxes Employee Benefits Educational Encollment	Educational Financing Health and Disability Work Related Expenses Training Questions (ETA)
Sept. 86-Dec. 86				6	Assets Liabilities	Open
Jan. 87-Apr. 87				7	Health and Disability Work History Education History	Open
May 87-Aug. 87				8	Annual Income Taxes Employee Benefits Education Enrollment	Educational Financing Training Questions (ETA)

Figure 5. SIPP Noninterview Data: Wave 1 and 2

	WAVE 1				
	Rotation 1	Rotation 2	Rotation 3	Rotation 4	Total Wave 1
Total Sample Size	6,465	6,420	6,403	6,626	25,914
Type B Noninterview	825 (12.8%)	832 (13.0%)	791 (12.4%)	904 (13.6%)	3,352 (12.9%)
Type C-W1 Noninterview	392 (6.1%)	393 (6.1%)	415 (6.5%)	443 (6.7%)	1,643 (6.3%)
Net Eligible for Interview	5,248	5,195	5,197	5,279	20,919
Type A Noninterview Total	267 (5.1%)	222 (4.3%)	272 (5.2%)	253 (4.8%)	1,014 (4.8%)
Refusal	194 (3.7%)	182 (2.5%)	197 (3.8%)	201 (3.8%)	774 (3.7%)
Other	73 (1.4%)	40 (.8%)	75 (1.4%)	52 (1.0%)	240 (1.1%)
Completed Interview	4,981	4,973	4,925	5,026	19,905

	WAVE 2			
	Rotation 1	Rotation 2	Rotation 3	Total Wave 2
Workload	5,120	5,144	5,116	15,380
Type C-W2 Noninterview	28 (.5%)	46 (.9%)	41 (.8%)	115 (.7%)
Type D Noninterview	57 (1.2%)	37 (.7%)	52 (1.0%)	146 (.9%)
Type A Noninterview Total	195 (3.9%)	185 (3.7%)	185 (3.7%)	565 (3.7%)
Refusal	157 (3.1%)	147 (2.9%)	160 (3.2%)	464 (3.1%)
Other	38 (.8%)	38 (.8%)	25 (.4%)	101 (.6%)
Completed Interview	4,840	4,876	4,838	14,554

Type A - Noninterview
- Refusal
- No one home
- Temporarily absent
- Unable to locate
- Other - Specify

Type B - Noninterview
- Vacant
- Occupied by persons with URE
- Unfit or to be demolished
- Under construction, not ready
- Converted to temporary business or storage
- Unoccupied tent site or trailer site
- Permit granted, construction not started
- Other - Specify

Type C-W1 Noninterview
- Demolished
- House or trailer moved
- Converted to permanent business or storage
- Merged
- Condemned
- Other - Specify

Type C-W2 Noninterview
- Entire household out-of-scope (For example: entire household deceased, institutionalized, etc.)

Type D - Noninterview
- Moved, address unknown
- Moved within country beyond limit

$$\text{Type A Noninterview Rate} = \frac{\text{Type A Noninterviews}}{\text{Interviews + Type A Noninterview}}$$

$$\text{Type B Noninterview Rate} = \frac{\text{Type B Noninterviews}}{\text{Work Load}}$$

$$\text{Type C Noninterview Rate} = \frac{\text{Type C Noninterviews}}{\text{Work Load}}$$

$$\text{Type D Noninterview Rate} = \frac{\text{Type D Noninterviews}}{\text{Work Load}}$$

Survey of Income and Program Participation

Papers Presented in the
Survey of Income and Program Participation
Session III
at the Annual Meeting of the
American Statistical Association
in Philadelphia, PA
August 13-16, 1984

No. 8406

U.S. BUREAU OF THE CENSUS
John G. Keane, Director
C.L. Kincannon, Deputy Director
William P. Butz, Associate Director
for Demographic Fields

POPULATION DIVISION
Roger A. Herriot, Chief

Acknowledgements

This publication is composed of papers prepared by many different authors for presentation at the American Statistical Association on August 13-16, 1984. We would like to thank these authors for their cooperation in making the papers available for publication. Clerical and editorial assistance was provided by Hazel Beaton, Mary Kisner, and Delma Frankel.

Suggested Citation

"Papers Presented at the Survey of Income and Program Participation session III at the annual meeting of the American Statistical Association in Philadelphia, PA, August 13-16, 1984," SIPP Working Paper Series No. 8406. U.S. Bureau of the Census, Washington, D.C. 1984.

Preface

This report is comprised of five papers featured in the "Survey of Income and Program Participation" session III, one of two in the Survey Research Methods Section of the annual meeting of the American Statistical Association.

This session covered a range of topics, both methodological and substantive, about longitudinal surveys and the Survey of Income and Program Participation (SIPP).

SIPP is a new Census Bureau survey collecting data that will help measure income distribution and poverty throughout the country more accurately. These data will be used to study Federal and state aid programs (such as food stamps, welfare, Medicaid, and subsidized housing), to estimate future program costs and coverage, and to assess the effects of proposed changes in program eligibility rules or benefit levels.

Households in the survey will be interviewed at 4-month intervals over a period of 2 1/2 years. The reference period will be the 4 months preceding the interview. In all, about 20,000 households will be interviewed, approximately 5,000 each month. Field operations will be handled through our 12 regional offices.

Recurring questions will deal with employment, types of income, and noncash benefits. Periodic questions will be added dealing with school enrollment, marital history, migration, disability, and other topics. Special supplemental questions will also be added to the SIPP questionnaire.

These papers discuss SIPP and its predecessor, the Income Survey Development Program (ISDP), an experimental program designed to test procedures used in conducting SIPP.

Session: Survey of Income and Program Participation III

Session Chair: Daniel Horvitz
Research Triangle Institute

Papers:

"Obtaining a Cross-Sectional Estimate from a Longitudinal Survey: Experiences of the ISDP."
Written by H. Huang, Bureau of the Census.

(Examines alternative cross-sectional weighting schemes.)

"Weighting of Persons for SIPP Longitudinal Tabulations."
Written by L. Ernst, D. Hubble, D. Judkins, D. B. McMillen, and R. Singh, Bureau of the Census.

(Discusses changes in the sample during the survey and various weighting methods.)

"Longitudinal Family and Household Estimation in SIPP."
Written by L. Ernst, D. Hubble, and D. Judkins, Bureau of the Census.

(Compares weighting approaches for use with longitudinal household and family concepts.)

"Early Indications of Item Nonresponse in SIPP."
Written by J. Coder and A. Feldman, Bureau of the Census.

(Studies nonresponse rates in the first SIPP interviews.)

Obtaining Cross-sectional Estimates from a Longitudinal Survey: Experiences of the Income Survey Development Program

by

Hertz Huang

Bureau of the Census

Paper to be presented at the 1984 meetings of the American Statistical Association, August 13-16, 1984.

Obtaining Cross-sectional Estimates from a Longitudinal Survey: Experiences of the Income Survey Development Program

I. INTRODUCTION

In 1975 the Secretary of the Department of Health, Education and Welfare (The Department of Health and Human Services (HHS) predecessor agency) authorized a program, the Income Survey Development Program (ISDP), to resolve technical and operational issues for a major new survey — the Survey of Income and Program Participation (SIPP). Much of the work of the ISDP centered around four experimental field tests that were conducted in collaboration with the Bureau of the Census to examine different concepts, procedures, questionnaires, recall periods, etc. Two of the tests were restricted to a small number of geographic sites; the other two were nationwide. Of the two nationwide tests, the more important data collection was called the 1979 Research Panel. This panel consisted of nationally representative samples which provided a vehicle for feasibility tests and controlled experiments of alternative design features. Information concerning the ISDP may be found in Ycas and Lininger (1981), David (1983), and the survey documentation now available through the National Technical Information Service (1983).

The 1979 Research Panel was a multiple frame sample consisting of a general population (area) sample of 9300 initially designated addresses drawn from the 1976 Survey of Income and Education (SIE) and some Census

Bureau's current survey reserve measures and new construction update, and two list frame samples of (a) eligible applicants for the Basic Educational Opportunity Grant (BEOG) Program and (b) blind and disabled Supplemental Security Income (SSI) recipients.

The 1979 Research Panel was a longitudinal survey consisting of six waves of interviewing; one third of the panel was interviewed each month with subsequent interview for a given unit occurring every three months. A sample of addresses was chosen and persons living in the sample units (addresses) during the first wave of interviews were defined as original sample persons. For interviews subsequent to the first, the sample of addresses became a sample of persons; accordingly, original sample people were followed to their new addresses in subsequent interviews (with reasonable geographic constraints — within 50 miles of any ISDP Primary Sampling Unit). Personal interviews were conducted in Wave 1 with all adults (persons sixteen years old and over) at the sampled address. These become the original sample persons. During Waves 2-6 all persons currently residing with an original sample person were interviewed. This means, for example, that if an original sample person moved to a new address with four other adults, then questionnaires were administered to everyone at the original sample person's new address. If any original sample person remained at the first wave address, anyone who moved into that address with the original sample person were also interviewed. Thus, interviews were conducted with all adults at an address as long as at least one of the adults present was an original sample person. Because of the ISDP rules, persons can be lost from

sample because they move beyond the survey's boundaries; in addition, people were added to the sample because they became part of the housing unit in which the original sample person resides.

Obviously, the universe changes continuously through the life of the survey. A great deal of interest exists, however, in developing cross-sectional estimates at the time of each interview wave. In the absence of drawing a new sample at each interview, any cross-sectional estimates developed for Waves 2-6 are subject to a population coverage bias. This paper will focus only on the covered population and present some unbiased base weights for cross-sectional estimators for the non-institutionalized U.S. population represented by the longitudinal sample (the population coverage bias will remain, however). Since the methodology for treating both area sample and list frame samples was needed for ISDP 1979 Research Panel, both will be described below. The estimation methods described here are directly applicable to the Survey of Income and Program Participation (SIPP), an overall description of which is found in Nelson, McMillen, and Kasprzyk (1984) and Herriot and Kasprzyk (1984).

II. THE POPULATION FOR CROSS-SECTIONAL ESTIMATES

We begin by defining the general population for which estimates are required. All households existing during the first wave of interviews (February through April 1979) are considered the initial population. Based on the rules adopted for the following individuals who move, we

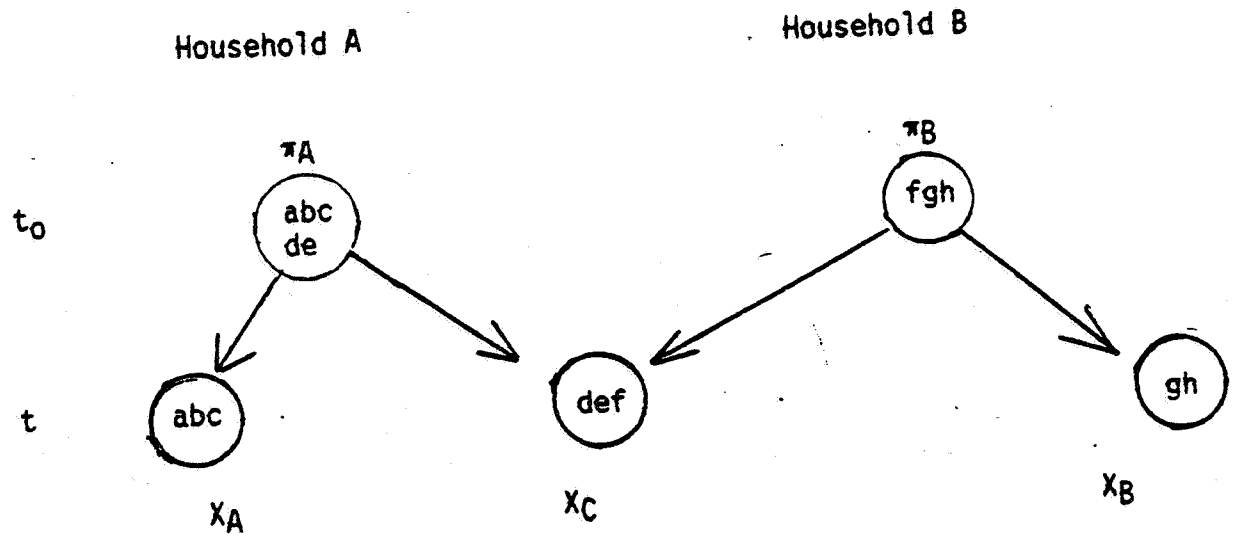
have essentially a longitudinal sample of persons as well as households for the initial population. Since no new sample was drawn at any subsequent interview, the sample does not completely represent the non-institutionalized U.S. population after first quarter of interview. There were persons in the following categories at the initial interview time but became part of the non-institutional population at a subsequent wave of interviewing: 1) U.S. citizens living abroad, 2) citizens of other countries who subsequently move to the U.S., 3) persons in institutions or armed forces barracks. These persons will be called the group R subpopulation which did not have chance to be selected as original sample persons. At a subsequent wave of interviews, the longitudinal sample did not include any household in which all current members were in the group R subpopulation. However, persons in the group R subpopulation who later joined households that included original persons eligible for sampling in the first wave were added to the cross-sectional universe. These person along with new borns will be called "additions" in subsequent waves. In general, "additions" are defined as persons moving into eligible households after the first wave who were not eligible for sampling in the first wave.

III. GENERAL CONCEPT OF CROSS-SECTIONAL ESTIMATION

Due to the procedures adopted for following movers in the 1979 Research Panel, at subsequent interviews a household could consist of members from more than one household in the universe at the time of the first

wave. The inclusion probability of such a household would depend on the inclusion probabilities of the households which the members of the current household were part of at the time of the first interview. The inverse of this inclusion probability is usually used as the weight of such a household in estimation. However, because of the sample design of the 1979 Research Panel, the inclusion probability of a household is a function of its primary sampling unit, type of sampling frames and the 1975 income of the household which occupied the housing unit during the SIE interview. Only the inclusion probability of an original sample household was feasible to calculate. The inclusion probability of an original nonsample household is almost impossible to evaluate. Therefore, some alternative weighting procedures needed to be explored.

The idea to be presented in this discussion is very simple. We will associate observations at any given point in time with the know inclusion probabilities of the original sample households. We will split up observations belonging to a household when current household members come from more than one original household. A portion of the observation is then associated with each original household. The following example will illustrate the idea: Assume that A & B are two original households with inclusion probabilities π_A and π_B respectively. At the first wave of interviews, household A consists of five members, a,b,c,d, and e, and the household B consists of three members, f,g, and h. During the second wave of interviews we find that d,e, and f are living together and form a new household, called household C, while a,b, and c are still in household A and g and h are still in household B.



Two alternatives are proposed, both involving the division of household C into two parts; one part is associated with household A and the other with household B:

a) Multiplicity Approach:

Based on the number of ways (called multiplicity) that the new household C can be included in the sample, the observation (additive, such as counts, income or values) of household C (called X_C) is divided by the number of original households involved (two in this case) and each portion is added to the corresponding observation of household A (called X_A) and household B (called X_B). Therefore, if both households A and B are original sample households, the cross-sectional estimate, $\hat{X}$, for the total at the second wave based on these three households can be expressed as:

$$\hat{X} = \frac{1}{\pi_A} (X_A + \frac{1}{2} X_C) + \frac{1}{\pi_B} (X_B + \frac{1}{2} X_C)$$

$$= \frac{1}{\pi_A} X_A + \frac{1}{\pi_B} X_B + (\frac{1}{2\pi_A} + \frac{1}{2\pi_B}) X_C.$$

Hence, the weight for the new household c is $\frac{1}{2\pi_A} + \frac{1}{2\pi_B}$. If only household A is a sample household, then the weight for the new household is $\frac{1}{2\pi_A}$; if only household B is a sample household then the weight for the new household is $\frac{1}{2\pi_B}$.

b) Fair Share Approach

This approach assumes that all household members contribute equally to their household. Thus, the observation of household C is divided into appropriate portions based on the proportion of members of household C which come from each original household (2/3 from household A and 1/3 from household B in this example). Therefore, if both households A and B are original sample households, the cross-sectional estimate for the total at the second wave based on these three households is expressed as

$$\hat{X} = \frac{1}{\pi_A}(X_A + \frac{2}{3}X_C) + \frac{1}{\pi_B}(X_B + \frac{1}{3}X_C)$$

$$= \frac{1}{\pi_A}X_A + \frac{1}{\pi_B}X_B + (\frac{2}{3\pi_A} + \frac{1}{3\pi_B})X_C.$$

Hence the weight for the new household C is $\frac{2}{3\pi_A} + \frac{1}{3\pi_B}$. If only household A is a sample household, then the weight for the new household is $\frac{2}{3\pi_A}$; if only household B is a sample household, then the weight for the new household is $\frac{1}{3\pi_B}$.

Since our sample was longitudinal in nature, if the universe remained constant through time, original sample persons would be a representative sample of the universe at any given point in time. Hence, using the inclusion probabilities of the original sample persons, the above estimators are unbiased (proof is given in next section). However, our feasible target population (excluding the group R subpopulation) changes through time by including the 'additions' (defined in Section II). To compensate for this, we will include these "additions" in the proposed estimators below.

IV. PROPOSED ESTIMATORS FOR GENERAL POPULATION (AREA) FRAME

Before the estimators are given, some notation should be defined. For the first wave of interview (time t_o), let

$X(t_o) = \sum_{k=1}^{N(t_o)} X_k(t_o)$ the parameter to be estimated, where $X_k(t_o)$ is the value of the characteristic for the kth unit (which may be a person or a household) and $N(t_o)$ is the number of units at time t_o;

$\alpha_k = 1$ if unit k was in the sample at time t_o, $k = 1, N(t_o)$

$= 0$ otherwise

π_k = the probability that unit k was selected for the sample at the first wave of interview (time t_o)

$= P_r(\alpha_k = 1) = E(\alpha_k)$, $k=1, N(t_o)$

At a subsequent wave (time t), define for each household i:

S_i = the total number of current residents of household i at time t

r_i = the number of original eligible households from which the current residents come (does not include households from which "additions" come)

and

$S_{i1}, S_{i2}, \ldots, S_{ir_i}$ be the number of current residents from each of r_i original households and S_{ia} be the number of current residents from the "additions" as defined in Section II. Write

$$S_i = \sum_{j=1}^{r_i} S_{ij} + S_{ia}$$

$$S_{io} = \sum_{j=1}^{r_i} S_{ij}$$

Also define N(t) as the total number of units in the target population at time t, such as household units (include all the original households plus newly formed households) or other units based on a group of persons such as families or sub-families (include all sample persons, interviewed nonsample persons plus "additions"). And let

$$X(t) = \sum_{i=1}^{N(t)} X_i(t)$$

be the parameter (total) to be estimated for the target population at time t. To simplify the notation, we will replace N(t), X(t) and $X_i(t)$ by N, X and X_i respectively.

Based on the general concept described in Section III, two cross-sectional estimators are proposed for the area frame to estimate the total of a characteristic of the target population at time t.

Estimator I (Multiplicity Estimator):

This estimator is based on the multiplicity of each current household

$$\hat{X}_M = \sum_{i=1}^{N} W_{Mi} X_i$$

where

$$W_{Mi} = \sum_{j=1}^{r_i} \frac{\alpha_j}{r_i \pi_j}.$$

Note that α_j and π_j are associated with original households but are reindexed within each current household i. It is easily seen that

$$E(\hat{X}_M) = E(\sum_{i=1}^{N} W_{mi} X_i) = E\left(\sum_{i=1}^{N} \sum_{j=1}^{r_i} \frac{\alpha_i}{r_i \pi_j} X_i\right)$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{r_i} \frac{E(\alpha_i)}{r_i \pi_j} X_i = \sum_{i=1}^{N} \frac{1}{r_i}\left(\sum_{j=1}^{r_i} \frac{\pi_j}{\pi_j}\right) X_i = \sum_{i=1}^{N} X_i = X$$

Therefore $\hat{X}_m$ is an unbiased estimator of X.

Estimator II (Fair Share Estimator):

This estimator is motivated by the assumption that all current household members contribute equally to the household in which they reside for the major household characteristic values, such as earnings and welfare benefits.

$$\hat{X}_F = \sum_{i=1}^{N} W_{Fi} X_i$$

where

$$W_{Fi} = \sum_{j=1}^{r_i} \frac{S_{ij}\,\alpha_j}{S_{io}\,\pi_j} .$$

As in the multiplicity estimator, α_j and π_j are associated with original households but are reindexed within each current household i. One can see that $\hat{X}_F$ is also an unbiased estimator of X as follows:

$$E(\hat{X}_F) = E\left(\sum_{i=1}^{N} W_{Fi} X_i\right) = E\left(\sum_{i=1}^{N} \sum_{j=1}^{r_i} \frac{S_{ij}\,\alpha_j}{S_{io}\,\pi_j} X_i\right)$$

$$= \sum_{i=1}^{N} \frac{1}{S_{io}} \left(\sum_{j=1}^{r_i} \frac{S_{ij} E(\alpha_j)}{\pi_j}\right) X_i = \sum_{i=1}^{N} \frac{1}{S_{io}} \left(\sum_{j=1}^{r_i} \frac{S_{ij} E(\alpha_j)}{\pi_j}\right) X_i$$

$$= \sum_{i=1}^{N} X_i = X$$

Note that if household j was not in sample at time t_o, it is unnecessary to know the number of current residents from original household j, S_{ij}, in $\hat{X}_F$ since $\alpha_j = 0$. Also note that because "additions" are not included in the weight calculations, they must be identified and excluded from using either estimator.

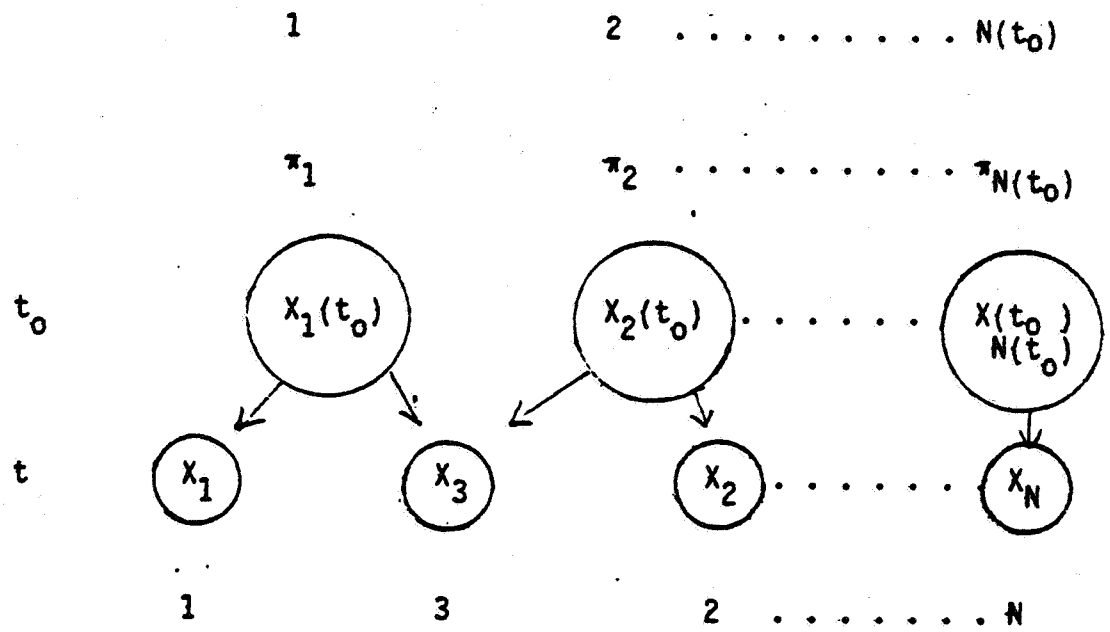
Comparison of Estimator I and Estimator II

Both Estimator I, $\hat{X}_M$, and Estimator II, $\hat{X}_F$, are feasible to compute. We now compare them in both areas of operations and reliability.

In order to compute $\hat{X}_M$, the number of original households eligible for sampling from which the current residents come is needed. This information is particularly difficult to obtain at each successive wave of the survey. However, to compute $\hat{X}_F$ one only needs to know the number of current residents in the household (excluding new additions) and the number or residents from each original sample household. This information could be obtained from the 1979 Research Panel person identifier without collecting additional information.

The equal contribution from the members of a household is a natural assumption. It reflects better the actual share among the household members in the absence of knowledge of the actual contribution from each member. For example, without knowledge of each person's age, employment status and other needed information, it is more logical to assume that earnings and welfare benefits are equally contributed by household members than any arbitrary way of defining household members' shares. And as will be seen below, this assumption also leads to the result that the estimator $\hat{X}_F$ has smaller variance than $\hat{X}_M$.

Assume that at a subsequent wave time t three households are generated from two original households of the first wave of interview (time t_o) as follows:



Let $X_k(t_o)$, k=1, $N(t_o)$ be the value of the characteristics of interest for household k at time t_o and X_j, j=1, N be that value for household j at time t. Using Section III we divide up X_3 in two parts, fX_3 and $(1-f)X_3$ and then associate fX_3 with X_1 and $(1-f)$ X_3 with X_2. Without loss of generality, assume that a sample size of 1 was selected at the first wave, to, with probability π_k, k=1, $N(t_o)$. An unbiased estimator of total, X, at time t can be written as

$$\hat{X} = \frac{\alpha_1}{\pi_1} X_1 + \frac{\alpha_2}{\pi_2} X_2 + \left(\frac{f\alpha_1}{\pi_1} + \frac{(1-f)\alpha_2}{\pi_2} \right) X_3 + \ldots\ldots\ldots\ldots\ldots$$

where α_i, $i=1$, $N(t_0)$ is defined at the beginning of this section. Notice that both $\hat{X}_M$ and $\hat{X}_F$ are special cases of $\hat{X}$. The variance of $\hat{X}$ is

$$Var(\hat{X}) = \pi_1 \left\{\left(\frac{1}{\pi_1} X_1 + \frac{f}{\pi_1} X_3\right) - X\right\}^2 + \pi_2 \left\{\left(\frac{1}{\pi_2} X_2 + \frac{1-f}{\pi_2} X_3\right) - X\right\}^2 + \ldots\ldots$$

The remaining terms are not explicitly given here since they are not functions of f. The $Var(\hat{X})$ is minimized if

$$f = \frac{\frac{X_2+X_3}{\pi_2} - \frac{X_1}{\pi_1}}{(\frac{1}{\pi_1} + \frac{1}{\pi_2})X_3}$$

Since usually not both π_1 and π_2 are known and in most of the surveys conducted by the Bureau of the Census, the inclusion probabilities, π_k, are about the same for all ultimate sampling units (even though they are unequal in the 1979 ISDP), one many simplify f to

$$f = \frac{X_2+X_3 - X_1}{2X_3}$$

Obviously, a weight defined as a function of survey observations is not easy to implement. To further simplify f, we assume the growth of X from t_0 to t is constant for all units and define

$$a\ X_1(t_o) = X_1 + X_{31}$$
$$a\ X_2(t_o) = X_2 + X_{32}$$
$$X_3 = X_{31} + X_{32}$$

where X_{3i} is the share of X_3 belonging to household members from original household i, i=1,2. Then

$$f = \frac{2\ X_{31} + a(X_2(t_o) - X_1(t_o))}{2(X_{31} + X_{32})}$$

Without knowledge of both $X_1(t_o)$ and $X_2(t_o)$, one would naturally assume that the two initial households are about the same i.e., $X_1(t_o) = X_2(t_o)$ and reduce f to

$$f = \frac{X_{31}}{X_{31} + X_{32}}$$

Now if the contribution to X_3 is proportional to the number of persons from each original household, then

$$f = \frac{S_{31}}{S_{30}}$$

as defined in W_{Fi}. This result can be extended to any sample size as well as that the new household members are from more than two original households. Therefore, without knowledge of the actual contribution from each household member, $Var(\hat{X}_F)$ is smaller than $Var(\hat{X}_M)$ under these general conditions.

V. PROPOSED ESTIMATORS FOR LIST FRAMES

Since persons are the list frame sampling units, we can divide all persons in the general population into three groups based on their relationship with the list frame under consideration.

I) Persons who are included in the list frame (called list frame persons). For the SSI list frame, this group includes all the nonaged recipients of the Federal Supplemental Security Income in December 1978; while for the BEOG list frame, this group includes all the eligible applicants of the Basic Educational Opportunity Grant as of September 1978 for school year 1978-79.

II) Persons who are not included in the list frame but live with a list frame person(s) during the first wave of interview (February through April 1979).

III) Persons who are not included in the list frame nor do they live with a list frame person(s) during the first wave of interview.

Both Group I and II had some chance to be included in the list frame sample, but Group III did not. The original (first quarter) households which consist of Group I and/or Group II persons will be called list frame households. As time went on, some members of Group III moved in and lived with person(s) belonging to Group I or II. Such members of Group III will be 'additions' for the list frame, since they are not initially eligible for sampling in the list frame. Note that the type of persons already described as "additions" for the general population (as defined in Section II) will also be "additions" for the list frame. For the following discussions, we now define two types of "additions" for the list frames: the "additions" that come from Group III will be called "Group III additions"and the type of "additions" as defined for the area frame will be called "area frame addition."

If a list of recipients of a government assistance program is used as a list frame then Group III is usually fairly large. If we construct our estimators the same way as we did for the area frame, we will include a lot of Group III persons in our estimates at time t of subsequent interviews. Consequently, we wouldn't really know what "subpopulation" we were estimating. In our opinion, it is not feasible to define such a subpopulation at time t. Without new sample drawn each wave from the updated list, a proper cross-sectional estimate for a list frame subpopulation at time t is not likely, especially if the turnover rate of the list frame members is high. Therefore, we will restrict our cross-sectional estimates to be based on only the original list frame sample persons (that is, the list frame persons selected for list frame sample plus all the persons who reside with them during the first quarter of interview) and the "area frame additions." In so doing we know that at any time t, the target population we are estimating consists of the original list frame subpopulation (that is Groups I and II) and the type of "additions" as defined in the area frame. Note that the original list frame subpopulation is determined by persons who were on the list at the time of sample selection. They may not be on the list by the time of initial interview.

In the 1979 ISDP panel, a household may have a multiple chance of being selected for the list frame sample if more than one member of the list frame persons live in that household at the first wave of interview.

(Some effort was made to reduce multiple chance of selection for those households in SSI frame.) Therefore, the concept of the base weight for the first wave of interview is no longer trivial.

Similar to the area frame, we define $X(t_o) = \sum_{i=1}^{N^L(t_o)} X_i(t_o)$ as the parameter to be estimated from a list frame sample at time t_o, where $X_i(t_o)$ is the value of the characteristic for the ith unit in the list frame subpopulation, which includes both Group I and II defined at the beginning of this section. Let

α_i = 1 if list frame person i is in the sample,

= 0 otherwise (note that $\alpha_i = 0$ for all non-list frame persons)

π_i = the probability that list frame person i is selected for the list frame sample for the first wave of interview (time t_o)

$= Pr(\alpha_i = 1) = E(\alpha_i)$

β_{oj} = the number of list frame persons (indexed by i) in the jth household at time t_o,

α'_j = 1 if the jth household is in the list frame sample,

= 0 otherwise.

Then the base weight at time t_o for the jth household and its residents is defined as

$$W_{oj} = \sum_{i=1}^{\beta_{oj}} \frac{\alpha_i}{\beta_{oj}\pi_i}$$

where α_i and π_i are associated with list frame persons but are reindexed within household j.

For time t of a subsequent wave, let

β_k = the total number of list frame persons living in the original (time t_o) list frame households which the current residents of the kth household come from.

S_k = the total number of current residents at time t; $S_{k1}, S_{k2}, \ldots, S_{kr_k}$ be the number of current residents in the kth household who come from each of r_k original list frame households; S_{ka} is the number of current residents of the kth household who are from the "area frame additions"; and $S_{k\ III}$ is the number of current residents of the kth household who are from the "Group III additions."

Therefore, $S_k = \sum_{j=1}^{r_k} S_{kj} + S_{k\ III} + S_{ka} = S_{kc} + S_{ka}$.

N^L = the total number of units such as household or family units, in the list frame universe at time t (note again that this includes both "area frame additions" and "Group III additions").

The two cross-sectional estimators for the total of a characteristic of the list frame target population at time t are as follows:

Estimator I (Multiplicity Estimator)

To avoid estimating "Group III additions" we will treat all the current residents from the "Group III additions" as a separate list frame sampling unit. Therefore, in the kth household at time t, there are $\beta_k + U_k$ list frame sampling units, where $U_k = 1$, if some of the current residents in the kth household are from "Group III additions," 0 otherwise. The multiplicity estimator for the list frame population total is given in the following:

$$\hat{X}_M^L = \sum_{k=1}^{N^L} W_{Mk}^L X_k$$

where

$$W_{Mk}^L = \sum_{i=1}^{\beta_k} \frac{\alpha_i}{(\beta_k + U_k)\pi_i}.$$

α_i and π_i are associated with original list frame person but are reindexed within each current household k.

Estimator II (Fair Share Estimator)

Motivated by the assumption that all current residents contribute equally to a household we propose the following list frame estimator:

$$\hat{X}_F^L = \sum_{k=1}^{N^L} W_{Fk}^L X_k$$

where

$$W_{Fk}^L = \sum_{j=1}^{r_k} \frac{S_{kj}\alpha_j'}{S_{kc}} W_{oj};$$

and α_j' and W_{oj} are associated with original household but are reindexed within each current household k.

These two estimators are constructed to estimate the list frame subpopulation excluding the "Group III addition." They are not unbiased estimators in global sense. However, the fair share estimator is unbiased under the assumptions that all current residents contribute equally to a household and a household is treated as a fraction of a household if some of the current residents are from "Group III additions." For example, suppose there are five persons, a,b,f,u and v in household M at time t. Among them, a and b are from original household A, f is from original household B and u and v are from "Group III additions." Furthermore, b and f are list frame persons. Denote X_i, i=a,b,f,u,v for the value of characteristic X of the ith person in the household. Then, the expected value of the fair share estimator of the characteristic X for household M is (3/5) $(X_a+X_b+X_f+X_u+X_v)$, and the corresponding value in our target population is $X_a+X_b+X_f$. Let $X_T = X_a+X_b+X_f+X_u+X_v$, then under the assumption of fair share estimator, $X_a+X_b+X_f = \frac{3}{5} X_T = \frac{3}{5} (X_a+ X_b+X_f+X_u+X_v)$. The expected number of households, for household M is $\frac{3}{5}$. For the multiplicity estimator, the situation is quite different. The expected value of X for household M is $\frac{2}{3} (X_a+X_b+X_f+X_u+X_v)$ and the expected number of households for household M is $\frac{2}{3}$. Therefore, more assumptions will have to be imposed before one can declare that the multiplicity estimator of X is unbiased.

In addition to the unbiasedness described above, $\hat{X}^L_F$ is also preferred for the same reasons (operational and reliability) stated in the area frame. In computing $\hat{X}^L_F$, we need to know β_{oj}, the number of list frame persons in a sample household at the initial interview (time t_o). This information was not difficult to obtain. And at any subsequent wave of interview time t, we needed to know only S_{kc}, the total number of current residents who are not "area frame additions" and S_{kj}, the number of current residents from each original list frame sample household. The latter can be obtained through the person identifier.

However, in order to compute $\hat{X}^L_M$ at time t we would have to ask one complicated question to obtain β_k, the total number of list frame persons living in the original households from which the current residents come. For example, for Supplemental Security Income (SSI) list frame at any subsequent interviews we would need to ask:

> During February through April 1979, how many of your household members were blind or disabled recipients of the Federal Supplemental Security Income in December 1978 who were neither institutionalized nor had representative payees?

It was determined to be extremely difficult to collect this information for the ISDP.

VI. SUMMARY

This research was completed before the first interviews of the 1979 ISDP. These two estimators were constructed based on the specific procedure of following movers in the 1979 ISDP. However, they can be easily modified to apply to other designs and procedures. The fair share estimator was actually used for the 1979 ISDP. It is also being used for the 1984 Survey of Income and Program Participation.

As noted in Section III, the inverse of the inclusion probability of a household at time t is usually used as the weight of that household to obtain unbiased estimator. When a household consists of members from two original households (called households i and j), the inclusion probability of this new household is $\pi_i + \pi_j - \pi_{ij}$, when π_{ij} is the joint selection probability of households i and j at the first wave of interview. This inclusion probability is not only operationally impossible to obtain, its inverse can be also reduced to the weight (W_{Mi}) of multiplicity estimator in most surveys conducted by the Census Bureau. In these surveys, the inclusion probabilities are almost the same for all ultimate sampling units and the joint selection probabilities are generally due to the sampling without replacement within PSU which are generally negligible. Therefore, the fair share estimator not only overcomes the trouble of obtaining such inclusion probabilities it is also a more reliable estimator than such traditional estimators under some general conditions and it is easy to implement.

SMD:HHuang:osh:jeg:LSB#2A

BIBLIOGRAPHY

[1] David, M. (ed), (1983). Technical, Conceptual and Administrative Lessons of the Income Survey Development Program. New York: Social Science Research Council.

[2] Herriot, R. and Kasprzyk, D. (1984). The Survey of Income and Program Participation. To be published in the Proceedings of The Social Statistics Section, American Statistical Associates.

[3] Nelson, D., McMillen, D. and Kasprzyk, D. (June 1984). An Overview of the Survey of Income and Program Participation. SIPP Working Paper Series, Number 84-01.

[4] Survey Documentation for 1979 Research Panel, (1983). Department of Commerce, National Technical Information Service, Springfield Virginia.

[5] U.S. Department of Commerce, Bureau of the Census, (1978). Some initial thoughts on Weighting the 1979 Income Survey Development Program. Memorandum from G. Capps to M. Schooley, November 30, 1978.

[6] U.S. Department of Commerce, Bureau of the Census, (1979). Some Alternative Procedures on Weighting the 1979 Income Survey Development Program (SDP). Memorandum from M. Schooley to G. Shapiro, March 1, 1979.

[7] Ycas, M. and Lininger, C., (1981). The Income Survey Development Program: Design Features and Initial Findings. Social Security Bulletin 44:11, November 1981.

WEIGHTING OF PERSONS FOR

SIPP LONGITUDINAL TABULATIONS

by

David R. Judkins, David L. Hubble,
James A. Dorsch, David B. McMillen
and Lawrence R. Ernst

U.S. Bureau of the Census

For Presentation at the American Statistical Association
Annual Meetings - August 1984

I. INTRODUCTION

Since October of 1983, the Census Bureau has been conducting interviews for a new survey, the Survey of Income and Program Participation (SIPP). The survey will effect long-sought improvements in the measurement of annual income and the complex relationships between income flows, labor force participation, participation in government programs such as welfare, and tax policy. One of the products of the interviewing will be a set of longitudinal records on a probability sample of the population. The subject we address in this paper is the weighting of these longitudinal records so that the data may be analyzed.

We are aware of only two precedents for this weighting. They are the National Medical Care Expenditure Survey (NMCES) and the National Medical Care Utilization and Expenditure Survey (NMCUES). The latter was conducted jointly by the Research Triangle Institute and the National Opinion Research Center [2]. Some work was done on the problem for the Income Survey Development Program (ISDP)[6], but it was not implemented. The techniques used by them are among those under consideration for SIPP. Naturally though, we are also considering some new ideas. These ideas are still in a very preliminary form. We are presenting them here to get early reaction and suggestions from the statistical community.

Our general approach consists of three major steps. The first step is to derive an unbiased weight for each longitudinal record. This is not as straightforward as it seems due to the fact that a slightly different set of people is being interviewed each month. Section III discusses this step.

The second step is to make adjustments for those records that are incomplete. We will use imputation when part of an interview is missing. (See Samuhel's paper in this session [3].) We will also probably use imputation when a

whole interview is missing where the missing interview is bracketed by good interviews. Our research on adjusting for records with more than one missing interview is in too preliminary a stage to report on. (One proposal has been made by Little and David [4].)

The third step is to correct for disproportional representation of demographic types to reduce variance and gain some consistency with the Current Population Survey (CPS). Section IV discusses this step.

Before discussing the weighting, it is essential that we define which of the many possible longitudinal universes is the universe for which estimates are to be provided. Section II deals with this problem.

Finally, we mention some of the important features of the design of SIPP. For more details, the reader is encouraged to first read an overview of the survey [5]. Roughly 20,000 households were interviewed between October 1983 and January 1984, inclusively. That set of interviews constitutes the first wave of the 1984 panel of SIPP. The Census Bureau will try to interview the persons in those households an additional seven or eight times in four-month waves, even if they move. We will also interview any persons who "usually reside" with anyone in the original cross-section for at least one-half of a calendar month. This extra interviewing will only be conducted for the time period that the joint residence is maintained. Only the original cross-section is followed through moves.

II. DEFINING THE LONGITUDINAL UNIVERSE OF PERSONS FOR SIPP

The SIPP universe at the beginning of any panel is persons who are members of the civilian non-institutional population, and members of the military not living in barracks on bases. Defining the longitudinal universe is somewhat more complicated. We begin by defining the possible ways persons can enter

and exit this universe. Next we discuss the relationship between the cross-sectional universes and the longitudinal universe. The third topic of this section addresses the definition of table universes, and a discussion of calculating annual income for persons in the longitudinal universe.

There are two ways persons can enter the SIPP universe: 1) persons can move from overseas (immigrate or return), institutions, or from military barracks; 2) persons can be born to members of the universe.

Similarly, there are two methods of exiting the universe; 1) moving overseas, to an institution, or to military barracks 2) dying. Given these conditions of entering and exiting the universe, and a definition of the initial universe, we can define the universe at any subsequent point in time, and the means by which the universe grows and diminishes over time. The next problem is to make the transition from the cross-sectional universes to a single longitudinal universe.

There are three methods of defining a longitudinal universe: 1) the composition can be fixed at some point in time; 2) the universe may be defined as the union of some set of cross-sectional universes; and 3) the universe may be defined as the intersection of some set of cross-sectional universes.

A longitudinal universe may be defined at a given point in time. For example, we can take the civilian noninstitutional population at the time the sample is drawn, at the midpoint of the panel duration, or at the end of the panel to define the universe of interest. Of course, the time point chosen could be any time point within the duration of the panel. This rather narrow

definition of the universe has an advantage in its simplicity, but also several disadvantages. Dependent on the chosen point in time, this definition produces a strictly declining population, a first increasing and then decreasing population, or a strictly increasing population. In the first case all entrants are excluded from the longitudinal universe, and only exits are allowed to alter the universe. In the second case, entry is allowed and exit is denied until the midpoint, when the situation reverses. In the last case, all those who exit during the panel are excluded from the longitudinal universe and only entries are allowed to alter the universe. In addition, it is difficult to argue why one point or another should be chosen as the point in time to define the universe, and for some purposes you may need a different point than the one originally chosen.

The next two definitions build from the above idea that a universe may be defined at any point during the panel. Let us assume then a set of universes each defined at a different point in time. To further simplify discussion, let us assume a set of twelve monthly universes defined at the midpoint of each month. The two options are to use either the union or the intersection of these sets.

Consider first the union of sets. The union of these monthly universes is all persons who were at some point during the year members of the civilian noninstitutional population. In other words, all members of the target population plus all persons who enter or exit during the year are included in the union of sets definition. This is the most inclusive of the universe definitions offered here, and the one which best captures the dynamic characteristics of the population. Some of the disadvantages of this type of definition will be raised below in the discussion of tabulations and

table universes.

An alternative to the union of sets is the intersection of the set of twelve monthly cross-sectional universes. Here we include in the longitudinal universe only those persons who were members of all of the cross-sectional universes. In other words, only those persons who were members of the civilian noninstitutional population or the special military categories on the fifteenth of each of the twelve months. This definition is even more restricted than the point-in-time definition. This intersection of sets definition produces a static population. That is to say there is no entering or exiting allowed.

Of the three longitudinal definitions offered here, only the union of sets incorporates the dynamic qualities that are inherent in a longitudinal process.

That would seem to make it the logical choice; however, this is also the definition that produces the most complications when tabulating data. These and other problems associated with tables are discussed below.

Consider, for example, a simple table of marital status given in figure 1. Here we are tabulating marital status at the beginning of the year with marital status at the end of the year. Thus, cell 1, 1 consists of those persons who were married at both points in time. Cell 1, 2 consists of those persons who were married at the beginning of the year and separated at the end of the year. Given the union of sets definition, there are not sufficient columns to tabulate all persons. In fact, nearly any universe definition will require the addition of at least one column to this table. That is to say, there is no place in this table for persons who were in

the universe at one point in time, and not in the universe at the other point in time. For the union of sets definition there is a need for both a column and a row for persons not in the universe at time 1 or not in the universe at time 2. For those definitions that allow exiting only a column for persons not in the universe at time 2 is necessary as long as the beginning point of the universe and the table are the same.

Similar problems arise in computing annual income. Aggregating across months is simple, but it is not clear how to compare income amounts for full year and part year persons. That is simply to say that a $6,000 income for 6 months and a $6,000 income for 12 months are not the same.

Figure 1

Marital Status 1984	Marital Status 1985				
	Married	Separated	Divorced	Widowed	Never Married
Married					
Separated					
Divorced					
Widowed					
Never Married					

III. INITIAL WEIGHTING

For SIPP, as for ISDP, a cross-section of the population will be followed for a period of time. Data will also be collected on the people that the original cross-section live with. The original idea was that only the data on the people in the original cross-section would be used in person longitudinal tabulations; the data on the other people would be used only to provide the "household experience" of the original cross-section. We are now reexamining that idea. The data on the other people can be used to better understand the experience of new entrants to the SIPP universe. Furthermore, there are ways to use these data more intensively to gain valuable variance reductions. Unfortunately, these procedures require strong assumptions for unbiasedness. In the following sections, we explore the trade-off. We begin with a general discussion, follow with a study plan on the question and some proposed procedures, and close with some examples on the application of the procedures.

A. Variance Reduction Versus Bias Control.

Let us first define some terms and clarify the type of parameters to be estimated. We divide the sample people into three groups. A person is an original sample person if he/she is a member of the original cross-section. 1/ A person is an associated sample person if he/she was a member of the eligible population at the time the cross-section was selected but happened not to be selected. Anyone else is an additional

1/ A person in original cross-section of households who was 15 years old or older at the time of the first interview is definitely an original sample person. Twelve, thirteen, and fourteen year old children are more difficult to classify. At first, no questionnaires are filled out for them and they are not followed in the rare event of an unaccompanied move. However, after they turn 15, they are treated the same as any other original sample person. We will treat them here as original sample people. Children eleven or younger are not classified at all.

sample person. This last group consists of recent discharges from institutions, new immigrants, and people moving out of military barracks. The type of parameter to be estimated is the frequency of some pattern of labor force participation, program participation, income receipt, etcetera, by demographic characteristics, housing characteristic, geographical unit, educational background, etcetera. A simple example is the frequency of women who were receiving public assistance in January 1984 but were not receiving it in December 1984.

The original idea was to estimate parameters like this one by summing the weights of all original sample persons with the desired characteristics. Data on associated and additional sample people are needed only to classify original sample people with respect to household characteristics; for example, was the original sample person living in a household in which at least one member received social security? Given this scheme, no data are needed on associated or additional sample people for the period that they don't reside with original sample people. Hence, we do not follow associated or additional sample people if they separate from original sample people. Clearly then, the data on associated and additional sample people are frequently incomplete.

Despite this incompleteness, we are now considering ways to squeeze more information out of this data. The first way is to provide estimates for the "union" universe using the data on additional sample people. The second way is to use the data on both types to reduce variances. To

begin the argument for this second use, we first point out that for shorter time periods these data are frequently either complete or nonexistent. (Throughout this section, by complete we mean complete ignoring nonresponse.) This is always true for 1 month periods, usually true for 3 month periods, and frequently true for 12 month periods. For example, suppose that Ruth is an original sample person interviewed in October 1983. In November, she marries Jack, who was in the October SIPP universe. They stay together at least through April 1985. Then Jack is an associated sample person on whom we have complete 1984 data. Alternatively, suppose that Jack was living in a military barracks in October 1983. Then he is an additional sample person on whom we have complete 1984 data. There will obviously be many more cases in these complete categories for 1985 data. Furthermore, there will be many cases where we are only missing one or two months of data.

Intuitively, it seems wasteful to give zero weights to these cases with complete or almost complete data, as originally intended. On the other hand, zero weights must be assigned to the seriously incomplete cases to avoid large-scale imputation. One possible solution is obtained by initally assigning strictly positive weights to all cases, including these that are incomplete due to field procedures, and then treating the incomplete cases as if they were caused by non-response. Imputation would be used for the almost complete cases, and a weighting adjustment would be used for the seriously incomplete cases. Note then that the seriously incomplete cases would have zero weights, while the other cases would have positive weights. If enough data has been collected on the associated and additional sample people to correctly model the probability of this type of nonresponse, then we would still have unbiased estimators.

An example the type of model required is that starting from a given socio-economic stratum, the new economic situation of a male divorcee does not depend on whether he or his ex-spouse was the original sample person. Here we stress that if a person has responded to even a single wave of SIPP, then we have an extraordinary wealth of data available for modeling.

Study Plan

Of course, we will never know for certain whether such a model is correct. There is a risk of biasing the estimators, and as a rule the Bureau is willing to risk biases for decreases in variance only if there is some evidence that the bias squared is substantially less than the variance decrease. Our plan at this time is to quantify for each proposed weighting procedure the frequency of positively weighted incomplete cases by the severity of the incompleteness. We will then have to rely on subjective judgement to determine if the variance decrease from aggressive use of incomplete data is worth the increased risk of bias. The only source for this information is the ISDP. We are currently working on ways to get appropriate tabulations for it.

B. Unbiased Weighting Procedures

Below we present a very simple result that characterizes a general class of unbiased procedures. Reflection on this result quickly helps one to understand that there are infinitely many unbiased procedures. Most of them are totally inappropriate, but it is very possible that better and radically different weighting procedures exist than have yet been conceived.

Let $X = \sum_i^N x_i$ be the parameter of interest to be estimated where x_i is the value of the characteristic for the ith unit. Let y_i be an unbiased estimator of x_i. Let w_i be a random variable associated with the ith unit such that y_i and w_i are independent and $E(w_i) = 1$.

Then $Y = \sum_i^N w_i y_i$ is an unbiased estimator of X because

$$E(Y) = E(\sum_i^N w_i y_i) = \sum_i^N E(w_i)E(y_i) = \sum_i^N x_i = X.$$

If we had complete response by all units, we could take $y_i = x_i$ and

$$w_i = \begin{cases} \text{inverse probability of selection if } i^{th} \text{ unit is in sample;} \\ 0 \text{ otherwise.} \end{cases}$$

Quite frequently, however, we will have incomplete response and will take y_i to be some imputation. Note that we may have $w_i = 0$ for units that are in sample, but that this may be disadvantageous because it wastes data.

Before we present the unbiased weighting procedures, let us define a term: cross-sectional person weight. The cross-sectional weight for a person is the cross-sectional weight of the household, of which he/she is a member. So, by defining the cross-sectional household weight we are implicitly defining the cross-sectional person weight. For simplicity, assume that the first wave cross-sectional weight for a sample household is the reciprocal of the probability of selection. For all nonsample households in the universe, this weight is zero.

Some household compositions may change during the period between two waves. For these sample households, new weights may have to be calculated to account for the changes. Using a multiplicity estimator of the general type suggested by Sirken [7], the cross-sectional house-

hold weight for any month after the first wave is defined to be the mean of the first wave cross-sectional household weights for all original and associated sample persons residing in the household that month.

In this section we present four longitudinal weighting procedures for computing unbiased estimates for persons. They are all presented in terms of the "union" universe, but they can be easily modified for the "intersection" universe by assigning a zero weight to any person who is not in every one of the 12 cross-sectional universes. In Section III.C we compare the procedures with respect to the use of data collected on associated sample persons and additional sample persons.

Procedure 1. Entry Date Weight (ED)

Each person receives a single longitudinal weight for any time interval that contains at least part of the period for which the person was in the universe, namely the cross-sectional weight for the person at his/her entry date into the universe. For all original and associated sample persons, the entry date into the universe is the start of the panel, so their longitudinal weights are their Wave 1 cross-sectional weights. For those who enter the universe after Wave 1, (additional sample persons), the longitudinal weight is the cross-sectional weight of the household, of which they are a member, as of the date they enter the universe. If the cross-sectional weight of the household at that date is zero, then the additional sample person's longitudinal weight is zero.

Procedure 2. Beginning Date of Time Interval Weight (BDI)

Each person receives a longitudinal weight valid for all time

Procedure 3. "Mid" Date of the Time Interval Weight (MDI)

This procedure is similar to Procedure 2. Each person receives a longitudinal weight valid for a specific time interval. Persons in the universe at the "mid" date of the time interval are assigned their respective cross-sectional weights at that date. The difference is that instead of the person longitudinal weights being determined at the beginning date of the time interval, these weights are determined at some predesignated date within the time interval. Persons that enter the universe during the time interval but after the mid date are assigned their respective cross-sectional weights as of the date they enter it, as in Procedure 1 and 2. Persons who leave the universe before the "mid" date are assigned their respective cross-sectional weights as of the date they leave it.

Procedure 4. Average Cross-Sectional Weight (ACS)

Each person receives a longitudinal weight valid for a specific time interval. Persons that remain in the universe throughout the interval are assigned the average of their respective monthly cross-sectional weights. Persons that enter or leave the universe are assigned the average of their respective monthly cross-sectional weights for the months they were in the universe during the time interval. Positive weights are assigned to all sample persons. A more formal definition is given below.

Let U_i = number of months the ith person was in the universe during the specified time interval

Let C_i = sum of the monthly cross-sectional weights of the ith person in the specified time interval

Then the person longitudinal weight is C_i/U_i.

C. Comparison of Procedures

In this section we describe in detail the types of complete and incomplete cases that are used by each procedure. First, we need to define some notation. Let

t_B = the first month that a person is in the universe,
t_E = the last month that a person is in the universe,

t_1 = the first month that a person is in sample,
t_2 = the last month that a person is in sample,

t_m = the mid-month of the interval of interest.

The description is given in Table 1. The first 14 cases comprise the "intersection" universe. The remaining 32 cases fill out the "union" universe. Each case is marked as having complete, partial or no data for the interval of interest. Of course, all of this is assuming perfect response. The only type of missingness that we are discussing here is that caused by operational procedures. On the right, there is a column for each procedure with an "X" if the procedure uses the case.

The entry date procedure uses the perfect cases 1,15,17, and 18, but does not use the perfect cases 2 and 16; the partial cases 3,5, and 19-27; and cases 12 and 44 for which no relevant data exists. The beginning date of interval and mid date of interval procedures both use all of the perfect cases, more of the partial cases and none of the completely missing cases. We thus think that these two procedures will tend to yield smaller variances than the entry date procedure with possibly some small increase in the risk of bias. The average cross-sectional procedure is the most aggressive in utilizing partial data. It uses all the perfect and partial cases and none of the completely missing cases. Also note that it assigns smaller

Table 1. Case Utilization by Procedure

Case	Preceeding Time Interval	Interval of Interest	Succeeding Time Interval	Completeness	Procedure ED	BDI	MDI	ACS
1	$t_B=t_1$		$t_2 \le t_E$	Perfect	X	X	X	X
2	$t_B<t_1$		$t_2 \le t_E$	"		X	X	X
3	$t_B=t_1$	$t_m \le t_2$	t_E	Partial	X	X	X	X
4	$t_B<t_1$	$t_m \le t_2$	t_E	"		X	X	X
5	$t_B=t_1$	$t_2<t_m$	t_E	"	X	X		X
6	$t_B<t_1$	$t_2<t_m$	t_E	"		X		X
7	t_B	$t_1 \le t_m$	$t_2 \le t_E$	"			X	X
8	t_B	$t_1 \le t_m \le t_2$	t_E	"			X	X
9	t_B	$t_1 \le t_2 < t_m$	t_E	"				X
10	t_B	$t_m < t_1 \le t_2$	t_E	"				X
11	t_B	$t_m < t_1$	$t_2 \le t_E$	"				X
12	$t_B=t_1 \le t_2$		t_E	No Data	X			
13	$t_B<t_1 \le t_2$		t_E	"				
14	t_B		$t_1 \le t_2 \le t_E$	"				
15	$t_B=t_1$	$t_2=t_E$		Perfect	X	X	X	X
16	$t_B<t_1$	$t_2=t_E$		"		X	X	X
17		$t_B=t_1$	$t_2 \le t_E$	"	X	X	X	X
18		$t_B=t_1$ and $t_2=t_E$		"	X	X	X	X
19	$t_B=t_1$	$t_m \le t_2 < t_E$		Partial	X	X	X	X
20		$t_B=t_1$	$t_2<t_E$	"	X	X	X	X
21		$t_B=t_1$ and $t_m \le t_2$	t_E	"	X	X	X	X
22		$t_B=t_1$ and $t_m \le t_2 < t_E$		"	X	X	X	X
23	$t_B=t_1$	$t_2<t_m \le t_E$		"	X	X		X
24	$t_B=t_1$	$t_2<t_E<t_m$		"	X	X		X
25		$t_B=t_1 \le t_2 < t_m$	t_E	"	X	X		X
26		$t_B=t_1 \le t_2 < t_m \le t_E$		"	X	X		X
27		$t_B=t_1 \le t_2 < t_E \le t_m$		"	X	X		X
28	$t_B<t_1$	$t_m \le t_2 < t_E$		"		X	X	X
29	$t_B<t_1$	$t_2<t_m \le t_E$		"		X		X
30	$t_B<t_1$	$t_2<t_E \le t_m$		"		X		X

Case	Preceeding Time Interval	Interval of Interest	Succeeding Time Interval	Completeness	Procedure ED	BDI	MDI	ACS
31	t_B	$t_1 \le t_m \le t_2 < t_E$		Partial			X	X
32	t_B	$t_1 \le t_m$ and $t_2 = t_E$		"			X	X
33		$t_B < t_1 \le t_m$	$t_2 \le t_E$	"			X	X
34		$t_B < t_1 \le t_m \le t_2$	t_E	"			X	X
35		$t_B < t_1 \le t_m \le t_2 \le t_E$		"			X	X
36	t_B	$t_1 \le t_2 < t_m \le t_E$		"				X
37	t_B	$t_1 \le t_2 < t_E \le t_m$		"				X
38	t_B	$t_m < t_1 \le t_2 \le t_E$		"				X
39		$t_B < t_1 \le t_2 < t_m$	t_E	"				X
40		$t_m < t_B < t_1$	$t_2 \le t_E$	"				X
41		$t_m < t_B < t_1 \le t_2$	t_E	"				X
42		$t_B < t_1 \le t_2 < t_E \le t_m$		"				X
43		$t_m \le t_B < t_1 \le t_2 \le t_E$		"				X
44	$t_B = t_1 \le t_2$	t_E		No Data	X			
45	$t_B < t_1 \le t_2$	t_E		"				
46		t_B	$t_1 \le t_2 \le t_E$	"				

weights, in general, to the partial cases than the perfect cases. We think it will tend to yield the smallest variances with the greatest risk of bias.

D. Examples

In these examples a divorced mother, previously living alone, has one of her children (older then 14) and her widowed mother move into her house in December 1983. All three had been in separate households prior to that date. In March 1984, her widowed mother remarries and her new husband, who also had previously been living along, moves into the house at that time. In May 1984, the child leaves the house and moves into an apartment by himself. It is also given that the divorced mother was an original sample person with a weight of 3600, and the child was an original sample person with a weight of 7200, both from rotation group 1 which was interviewed in October 1983. Determine the longitudinal person weight for each of these four persons for the entire year 1984, for each of the procedures, with the following two scenarios:

1. All four persons were in the universe throughout the life of the sample.

2. The same, except now the widowed mother was in an institution until December 1983.

Let

D = divorced mother

C = child

W = widowed mother

H = husband

Entry Date Procedure

Scenario 1.

Since all four people were in the universe for the first wave, the weights are their first wave cross-sectional weights, that is

D = 3,600, C = 7,200, W = 0, H = 0.

Scenario 2.

The same, except the widowed mother's weight is now the cross-sectional weight of the household in which she was residing when she entered the universe in December, 1983, that is W = 5,400 = (3600+7200)/2.

For the other three procedures, we first compute the cross-sectional household weights associated with each person for every month of 1984, for both scenarios. The results are given in Table 2.

Table 2

Monthly Cross-Sectional Weights

	Scenario 1			Scenario 2		
	Jan.-Feb.	Mar.-Apr.	May-Dec.	Jan.-Feb.	Mar.-Apr.	May-Dec.
D	3,600	2,700	1,200	5,400	3,600	1,800
C	3,600	2,700	7,200	5,400	3,600	7,200
W	3,600	2,700	1,200	5,400	3,600	1,800
H	0	2,700	1,200	0	3,600	1,800

The reason for the differences in the weights between these two scenarios is that the widowed mother does not enter into the denominator in Scenario 2.

From the above table, the weights for the beginning date of time interval, and mid-date of time interval procedures immediately follow, while for the average cross-sectional weight procedure, we simply average over the twelve months. The results are given in Table 3. The weights from the entry date procedure are also shown.

Table 3

Longitudinal Weights

Procedure / Person	Scenario 1				Scenario 2			
	ED	BDI	MDI	ACS	ED	BDI	MDI	ACS
D	3,600	3,600	1,200	1,850	3,600	5,400	1,800	2,700
C	7,200	3,600	7,200	5,850	7,200	5,400	7,200	6,300
W	0	3,600	1,200	1,850	5,400	5,400	1,800	2,700
H	0	0	1,200	1,250	0	0	1,800	1,800

IV. CONTROLS

We are currently considering the adjustment of SIPP longitudinal weights so as to achieve the variance reductions associated with ratio estimation while also causing agreement with SIPP cross-sectional controls on a monthly basis; i.e., in addition to simple undercoverage adjustments we are considering the possibility of forcing the sum of the longitudinal weights of all persons in the universe in a given month to equal the cross-sectional population control for that month. Since longitudinal weights are fixed over time while the universe fluctuates over time, such agreement will not occur unless proper steps are taken to ensure it. We are also considering adjustments to force spouses to have equal longitudinal weights. We are considering these two possibilities in order to enhance the face validity of the survey at the least possible cost of reduced precision.

A. Objectives

The primary reason for ratio adjustment of longitudinal weights is to reduce variances of longitudinal weights by ensuring representativeness with respect to demographic variables which are highly correlated with

the variables to be measured. (This is frequently referred to as post-stratification.) To the extent that it corrects for differential under-coverage, it is also hoped that bias is reduced by ratio adjustment.

A reasonably good adjustment is to proportionately adjust the weights of persons by demographic type in a specified month so that the weighted counts agree with independent population estimates by demographic type for that month. Persons not in sample in the chosen month are assigned the factor for their demographic type. This approach operates under the assumption that the degree to which the sample represents each demographic type is not highly variable over time. This adjustment does not adjust weights to monthly controls other than those for the chosen month. Another approach is to make the adjustment for all persons for each of the 12 data months, then assign to a person the average of the 12 factors for his/her cell. Such an adjustment would tend to be influenced less by the vagaries of sample selection. Addressed here is the more complex problem of adjusting weights for disproportional representation in a manner such that consistency with cross-sectional controls is achieved for each month. This problem has a multitude of solutions. However, the solution we seek should be the one which provides the greatest variance reduction. One possible solution is to first adjust weights as outlined in the above paragraph, then further adjust them so that the desired monthly consistency is achieved while minimizing the amount by which weights are further adjusted. This approach preserves the benefits of the initial adjustment by demographic variables provided that this second adjustment causes relatively small changes in weights.

A further refinement would be to adjust so that spouses have equal weights. Naturally, persons undergo changes in marital status during the year; some persons may have more than one spouse over a one year period. Define a "marriage group" to be a group of persons in the SIPP sample, each of whom has been or is married to at least one other person in the group during the data year. It is possible to perform an adjustment so that all persons in a given marriage group have equal weights. This last adjustment would cause slight disagreements between longitudinal population estimates and monthly controls; we believe that such disagreements could be made arbitrarily small.

B. Outline of Adjustment Process

The basic steps in the adjustment process are as follows:

1. Post-stratification

The demographic types that we are most likely to use in post-stratification are those defined by age, race, sex, and household relationship. These are similar to the types used in post stratification for cross-sectional estimation [1].

Within each rotation of the SIPP sample, all persons would receive an adjustment factor ensuring representativeness of the types discussed above. Two possible adjustments are currently under consideration.

a. One month adjustment

For each rotation of the SIPP sample, an initial adjustment is performed for a single month. The weights of all sample persons in the rotation are adjusted so that, for each cell, (demographic type) the sum of that rotation's person longitudinal weights (through this stage of adjustment) is equal to 1/4

of the cell's cross-sectional control for the chosen month. (The factor 1/4 reflects the fact that only 1/4 of the SIPP sample is designated for interview in a given month.) There may be reasons to choose a particular month for the adjustment, but the month chosen has no effect on this development.

Having computed a post-stratification adjustment factor for each of the above defined cells, the factor for the appropriate cell is also applied to each person who is in the SIPP sample during some part of the data year but not in sample during the month for which the adjustment factors are computed.

b. Multiple month adjustment

A second approach is to compute adjustment factors as in a. for all 12 months of the data year, then average the 12 factors for each cell. Either a weighted or unweighted average might be used. This type of adjustment would tend to be smoother than a one month adjustment, and would likely require less adjustment in 2., immediately below.

2. Cross-sectional Consistency Adjustment

It has been proposed that some form of adjustment be used to cause consistency with cross-sectional controls during each month of the data year. If possible, (i.e., if there are enough sample persons to yield reliable adjustment factors), the adjustment would be performed by some small number of cells (perhaps 4 age x sex cells). For simplicity, a one cell adjustment is discussed here. Since adjustments prior to this one should offer substantial variance reductions, the object of the proposed adjustment would be to achieve

the desired consistency while minimizing the overall adjustment to the current weights. Two approaches to the problem are currently under consideration.

The following notation will be used in both developments. Let

Y_{ij} = the weighted number of persons in the SIPP sample only from month i through month j of the data year, where $i \leq j$. The weight is the adjusted weight from 1. above.

X_{ij} = the (unknown) weighted number of persons analogous to Y_{ij}, after the cross-sectional consistency adjustment.

C_k = the cross-sectional control for the kth month.

There are 78 Y_{ij}'s and 78 X_{ij}'s.

a. Lagrange multiplier approach

This approach seeks to minimize the sum of squared deviations

$$D_1 = \sum_{j=1}^{12} \sum_{i=1}^{j} W_{ij} \, (Y_{ij} - X_{ij})^2 \qquad (1)$$

subject to the constraints

$$\sum_{i=1}^{k} \sum_{j=k}^{12} X_{ij} = C_k, \; k=1, 2, 3, \ldots, 12, \qquad (2)$$

where W_{ij} is an arbitrar y weight. The problem can be easily solved with Lagrange multipliers.

After solving for X_{ij}, the cross-sectional consistency adjustment factor to be applied to the weight of each person in sample only from the ith through j^{th} months is set equal to X_{ij}/Y_{ij}.

Though this problem generally has a solution, it is possible

that negative or large positive weights will result.

b. Linear programming approach

This approach guarantees non-negative weights but may not provide perfect consistency. The idea here is to minimize

$$D_2 = \sum_{j=1}^{12} \sum_{i=1}^{j} |X_{ij} - Y_{ij}| \qquad (3)$$

subject to the constraints (2) and the additional constraints

$$X_{ij} \geq 0, \; i, j = 1, 2, 3, \ldots, 12 \qquad (4)$$

The objective function (3) can be expressed in a form by which the problem can be written as a linear programming problem. This problem, if it has a solution, results in non-negative weights due to the constraints (4). It is possible that the problem has no solution, in which case the constraints must be relaxed to the extent that a solution becomes possible. It is possible to write constraints which keep weights from becoming larger than some arbitrary value; it is possible that these additional constraints could make it necessary to relax other constraints. Alternatively, the constraints might be limited in number by requiring consistency for only a subset of the 12 months.

3. Marriage Group Equalization Adjustment

Recall that a marriage group is defined as a collection of persons, each of whom is or was married to another person in the group during some part of the time period over which longitudinal weights are computed. For consistency purposes, it is desirable that persons within a marriage group have equal longitudinal weights. This can be achieved while ensuring that weights will sum to within some specified tolerance of cross-sectional controls for each month,

using the following iterative procedure.

a. Each person within each marriage group is assigned the average weight of persons in that marriage group, using weights adjusted through the cross-sectional consistency adjustment.

b. One of the procedures (whichever is chosen for cross-sectional adjustment, or perhaps some other method) discussed in 2. is implemented, using weights adjusted for cross-sectional consistency to determine Y_{ij} and defining X_{ij} to be the weighted number of persons, analogous to Y_{ij}, after the current adjustment. This yields an adjustment factor to be applied to each person's weight.

c. Steps like a. and b. above can be carried out, one after another, continuing to use the most current weight. Each time step a. is repeated, a check is made to determine whether all Y_{ij}'s are within the specified tolerance of the respective Y_{ij}'s from 2.

d. When the tolerance is met or exceeded by each Y_{ij}, the procedure is terminated and final longitudinal weights are assigned as follows:

   i. Each person in a marriage group receives the last average weight computed for his/her marriage group.

   ii. Each remaining person is assigned a final longitudinal weight equal to his/her weight, through cross-sectional consistency adjustment, multiplied by the product of the factors computed at each successive operation b. above.

It has not been determined whether the above procedure would

necessarily converge, although convergence appears likely. A final remark is that neither of these last two described adjustments would be completely beneficial. They would cause some (to our belief small) increase in mean square error.

REFERENCES

[1] Census Bureau memorandum from C. Jones to T. Walsh, "Cross-Sectional Weigting Specifications for the First Wave of the 1984 Panel of the Survey of Income and Program Participation (SIPP)," November 25, 1983.

[2] Jones, Bruce L., "Development of Sample Weights for the National Household Survey Component of the National Medical Care Utilization and Medicare Utilization and Expenditure Survey," April 1982.

[3] Samuhel, Michael E., "Longitudinal Item Imputation in a Complex Survey," presented to the Survey Research Methods Section of the American Statistical Association during the 1984 Annual Meetings.

[4] Little, R.J.A. and David, M., "Weighting Adjustments for Nonresponse in Panel Surveys," 1983 Working Paper.

[5] Nelson, D., McMillen, D., and Kasprzyk, D., "An Overview of the Survey of Income and Program Participation," SIPP Working Paper Series No. 8401. U.S. Bureau of the Census, Washington, D.C. 1984.

[6] Kasprzyk, D. and Kalton, G., "Longitudinal Weighting in the Income Survey Development Program," in Technical, Conceptual and Administrative Lessons of the Income Survey Development Program (ISDP), Papers presented at a conference, October 6-7, 1982. Social Science Research Council, Washington D.C., 1983.

[7] Sirken, Monroe G., "Household Surveys with Multiplicity," Journal of the American Statistical Association, 65, No. 329 (1970), 257-66.

LONGITUDINAL FAMILY AND HOUSEHOLD ESTIMATION IN SIPP

by

Lawrence R. Ernst, David L. Hubble, and David R. Judkins
Bureau of the Census

For Presentation at the
American Statistical Association Annual Meeting
Philadelphia, Pennsylvania
August 1984

1. INTRODUCTION

Many types of statistics will be produced by the Survey of Income and Program Participation (SIPP), but there is one type that was the driving force behind the unique design of the survey. To be fully successful, SIPP must tell us what happens to households over the course of time. From it we must obtain estimates of the patterns of income receipt, program participation, and labor force participation at the household and family level by a host of other characteristics. Of particular interest are parameters such as total annual household income and the number of families that have stopped drawing food stamps by demographic characteristics.

Before estimates can be produced, a decision must be made on the definition of a longitudinal household to be used in this survey. (To simplify the presentation, we will concentrate our discussion on longitudinal households as opposed to longitudinal families. However, parallel longitudinal estimation procedures can readily be developed for families). It often happens that the occupants of several housing units move and regroup. We need to know which, if any, of the resulting households are to be considered continuations of the previous households. Many definitions have been proposed, but final agreement has thus far not been achieved. Also decisions have yet to be made on whether households that form or dissolve during a time interval of interest are to be considered as part of the universe for estimation purposes. Because of the absence of agreement in these areas, several proposed definition and universe combinations will be considered in this paper. They are listed in Section 2. Also because of this absence of agreement, the major aim of this paper will be simply to compare several possible longitudinal household estimation procedures and present criteria for choosing among them, without attempting to reach a conclusion on a preferred procedure.

We foresee several steps in the process of producing longitudinal households estimates. The focus in this paper, except for the final section, is the first step, the production of weights that would yield unbiased estimates assuming there are no data that are missing or in error, and that the frame coverage is perfect. Several procedures for obtaining such weights will be presented in Section 3. In Section 4 some numerical examples of the weights produced by these procedures are given. Choosing among these procedures is complicated by the fact that even assuming perfect response, data needed to produce unbiased estimates will be missing for some households because they are not collected with the current field procedures. This difficulty is principally due to the fact that, except for a few household definitions, all unbiased procedures assign positive weights to some longitudinal households for time periods when they are not in sample. The severity of this problem and the extent to which it is correctable in the future by changing field procedures or by modeling the missing data, vary by procedure. This problem, along with descriptions of other important features, both positive and negative, that estimation procedures may possess is presented in Section 5. This is followed in Section 6 by a detailed comparison of the features of the estimation procedures under consideration in this paper. Finally, in Section 7 we briefly discuss adjustments to the unbiased weights. It is anticipated that the two major components of such adjustments will be a procedure for adjusting for missing data, and a method for controlling key variables to independent estimates, such as CPS estimates.

It is assumed in this paper the reader has a basic knowledge of SIPP, including the design of this survey. Nelson, McMillen, and Kasprzyk (1984) provides this information.

2. LONGITUDINAL HOUSEHOLD DEFINITIONS

In this section four possible longitudinal household definitions are presented to illustrate the longitudinal weighting procedures that will be described in the next section. A thorough discussion of longitudinal household definitions is presented in McMillen and Herriot (1984). In addition, several other terms will be defined, including the longitudinal household universes considered in this paper.

Since household composition and data for SIPP are obtained on a monthly basis, each of the definitions to be presented will be in terms of household continuity from one month to the following month. A longitudinal household over a time interval of n (≥2) months is then defined to be one which is continuous for each of the n-1 corresponding pairs of consecutive months. (It has not yet been decided if this approach will actually be used in SIPP.)

For each of the definitions below the conditions for which household B at month t+1 is the continuation of household A at month t are stated. One condition that we require that all the definitions share is that A and B are either both family households or both non-family households. The other conditions are:

No Change Definition (NC). A and B have the same household members.

Same Householder (SH). A and B have the same householder. As an alternative, householder could be replaced by principal person in this definition without altering any of the statements made about it in subsequent sections, provided the final estimation procedure in Section 3 is also modified accordingly. (The householder of a household is, roughly, the person who owns or rents the housing unit. The principal person is the wife in a married-couple household, and the householder in all other households.)

Reciprocal Majority (RM). The majority of individuals who are both household members of A at time t and in the universe at time t+1 are members

of B at time t+1, and the majority of individuals who are both household members of B at time t+1 and in the universe at time t are members of A at time t. (This type of longitudinal definition was originally developed by Dicker and Casady (1982) for use in the National Medical Care Utilization and Expenditure Survey (NMCUES).)

Shared Experiences Definition (SE). Either conditions (1.a, b) or (2.a-e) presented below are satisfied.

(1.a) A and B are nonfamily households with the same householder, including single person households.

(b) At least 1/2 the members of A are members of B.

(2.a) A and B are family households.

(b) The householder or spouse of the householder of A is the householder or spouse of the householder of B.

(c) A and B have at least two members in common.

(d) If another household A' when substituted for A in (2.a-c) satisfies these conditions, then the number of household members common to A and B is more than the number common to A' and B.

(e) If another household B' when substituted for B in (2.a-c) satisfies these conditions, then the number of household members common to A and B is more than the number common to A and B'.

Some variation of this last definition is currently the leading candidate to be chosen as the SIPP longitudinal household definition.

We will now clarify several other terms.

A household is said to be in existence over a time interval of $n \geq 2$ months if it is longitudinal over that time interval. Its period of existence is the longest such time interval. In the case of a household which is defined cross-sectionally for a month t, but is not longitudinal over either of the two

month intervals containing t, then the period of existence of the household is defined to be one month.

If t_1 and t_2 are any pair of months, and longitudinal estimates are to be made over the interval $[t_1, t_2]$, then the following two possibilities will be considered in subsequent sections for the universe of households for which estimates will be produced.

Restricted Universe. The set of all households in existence over the entire interval $[t_1, t_2]$.

Unrestricted Universe. The set of all household in existence for one or more months in $[t_1, t_2]$.

Each sample panel is interviewed eight times. Each of the eight rounds of interviews takes four consecutive months to complete and is known as a wave.

Finally, we define an original sample person to be a person that was in sample during the first wave and will be at least 15 years of age by the end of the panel.

3. UNBIASED WEIGHTING PROCEDURES

In this section we present five weighting procedures for computing estimates of totals or proportions for longitudinal households that would be unbiased in the sense that the expected value of the estimator over all possible samples is the parameter of interest assuming no data are missing or in error, and perfect frame coverage. Modifications and adjustments of these estimation procedures necessary because of the unrealistic nature of these assumptions will be considered in Section 7. Except for the Continuous Household Members procedure, which will only be applied to the restricted universe, all the procedures will be stated for the unrestricted universe. To apply them to the restricted universe simply zero weight each household which is not in continuous existence over the time interval of interest.

Furthermore, unless otherwise stated, all the procedures will be applied to all four longitudinal definitions defined in Section 2.

First we will explain why a common method of estimation, weighting by the reciprocal of the probability of selection is not feasible for our purposes, and hence the need to consider alternative procedures. Let $X = \sum_{i=1}^{N} x_i$ be a parameter of interest, where x_i is the value of the characteristic for i-th unit in a population of size N. Typically in survey work, to estimate X a sample would be drawn in such a manner that the i-th unit has a known positive probability p_i of being chosen, and X would then be estimated by

$$\hat{X} = \sum_{i=1}^{N} w_i x_i \ , \tag{3.1}$$

where

$$w_i = \begin{cases} \dfrac{1}{p_i} & \text{if the i-th unit is in sample,} \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

Unfortunately for household and family estimation in SIPP, both cross-sectionally and longitudinally, such an estimation approach is not practical. For example, cross-sectionally a household is interviewed and used in the estimation process for a given month if and only if at least one household member is an original sample person. Consequently, to use (3.1) and (3.2) as an estimator it would be necessary to determine the probability that at least one member of the current household is an original sample person. It would be operationally impossible to determine this probability, since it would first be necessary to determine the first wave households for all current household members and then compute the probability that at least one of these first wave households was selected.

Fortunately though, it is not necessary that w_i satisfy (3.2) in order that (3.1) be unbiased. In fact if w_i is any random variable associated with the i-th unit in the population satisfying

$$E(w_i) = 1, \tag{3.3}$$

then (3.1) is unbiased, that is $E(\hat{X}) = X$. Thus, defining unbiased longitudinal household and family weighting procedures reduces to defining random variables w_i satisfying (3.3).

Before we present the longitudinal weighting procedures we will state what, for purposes of this paper, a cross-sectional household weight is, since most of longitudinal weighting procedures will be defined in terms of cross-sectional weights. The first wave cross-sectional weight for a sample household is taken here to be the reciprocal of the probability of selection. For all nonsample households in the universe this weight is defined to be zero. For any month after the first wave a different definition is necessary because of possible changes in household composition. So, the cross-sectional household weight for any such month is defined to be the mean of the first wave cross-sectional household weights for all persons in the household that month who will be at least 15 years of age by the end of the panel and who were in the universe during the first wave. This type of weighting procedure is currently being used in SIPP to produce cross-sectional estimates, hence the name. It is readily verifiable that the weights satisfy (3.3).

We also will leave it to the reader to verify that the weights for each of the longitudinal procedures to be presented satisfy (3.3) and hence lead to unbiased estimators.

<u>Beginning Date of Household Procedure (BH)</u>. Each longitudinal household receives a single weight valid for any time interval that contains at least part of the period for which the household existed, namely the cross-sectional

weight for the household at the beginning date of the household. In particular, if there were no original sample persons in a household at its beginning date then its longitudinal weight would be zero. This approach to longitudinal household estimation was previously used in the NMCUES (Whitmore, Cox and Folsom 1982).

Beginning Date of Time Interval Procedure (BI). Each longitudinal household receives a longitudinal weight valid for all time intervals with the same beginning date, namely the cross-sectional weight for the household at the beginning date of the time interval. Longitudinal households that form during the time interval are assigned the cross-sectional weight for the household at its beginning date, as in the preceeding procedure.

Continuous Household Members Procedure (CM). The following procedure will only be applied to the restricted universe, as defined in Section 2. For any time interval for which the household is in existence the longitudinal weight to be assigned is determined by the set of persons that are members of the household throughout the time interval. The longitudinal household weight is the cross-sectional weight that would be assigned to a household consisting of this set of persons; that is, the average of the first wave weights of these people. A longitudinal weight of zero is assigned to the household if there are no original sample persons who are members throughout the time interval. The procedure is slightly biased because a longitudinal household with no members continuously present throughout a time interval has no chance of receiving a positive weight, thereby making satisfaction of (3.3) impossible. Since we believe this situation will rarely occur, at least for the longitudinal household definitions considered here, we expect this bias to be very small.

Average Cross-Sectional Household Weight Procedure (AW). Each longitudinal household receives a longitudinal weight valid for a specific time

interval, namely the average of the monthly cross-sectional weights for the household over the intersection of the life of the household and the specified time interval.

Note, there are many procedures, like AW, that entail the averaging of weights, both household cross-sectional weights and person longitudinal weights. We will examine only one of these procedures here, as an example of this type of longitudinal household weighting procedure.

Householder Weight Procedure (HW). The following procedure will be applied only to the No Change and Same Householder Definitions, since it is appropriate only for definitions that allow for a single householder during the household's existence (Generalizations of this procedure which are not so restricted in their applicability exist but will not be considered here.) The procedure assigns a single weight valid for any time interval that contains at least part of the period for which the household existed, namely the first wave cross-sectional household weight of the householder's first wave household. A longitudinal weight of zero is assigned to the household if the householder was not an original sample person.

As will be seen in Section 5, this procedure is clearly the one of choice when the Same Householder Definition is used. If that type of definition is used with householder replaced by principal person then a similar modification of this estimation procedure with householder replaced by principal person would be appropriate.

4 . EXAMPLES

In the following examples, the cross-sectional weight for the second and subsequent waves will be as defined in Section 3. The longitudinal household definition for procedures BH, BI, CM, and AW will be the reciprocal majority rule, as given in Section 2. For procedure HW, the longitudinal household definition will be the same householder rule, as given in Section 2.

In these examples a divorced mother (householder) with two children (both older than 14) residing with her has her widowed mother move into her house in December, 1983. In August, 1984 her widowed mother remarries and the new husband moves into the house at that time. In April, 1985 one of the children leaves the household. The longitudinal household weights will be determined for the three procedures for the following time periods:

A. the entire year 1984;

B. the entire year 1985;

C. the entire years 1984-85.

This will be done in each case for the following two scenarios:

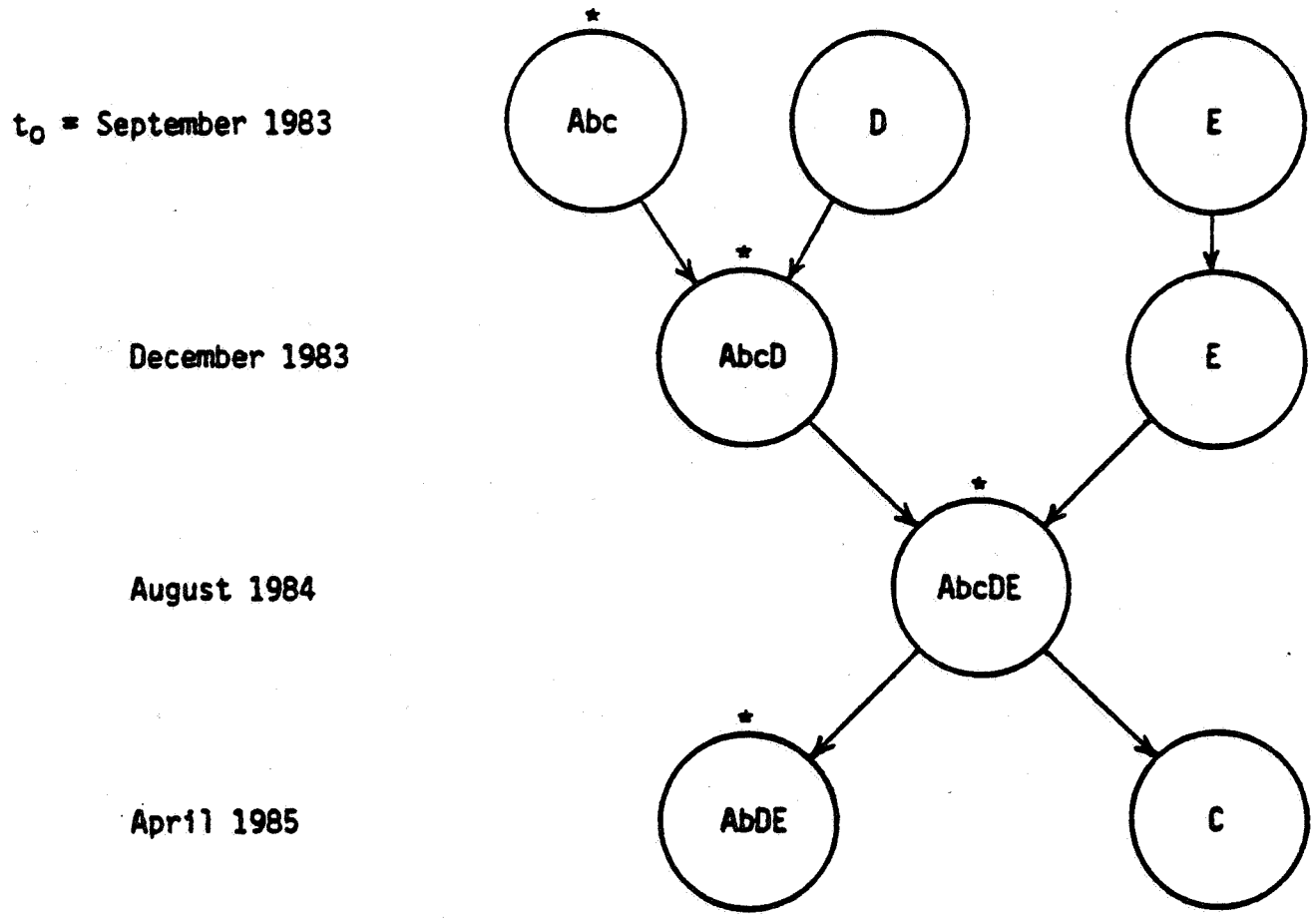
1. the new husband of the widowed mother was the only original sample person in the 1984 SIPP panel (originally interviewed in October, 1983-rotation group 1), with a first wave weight of 8,000;

2. in addition, the divorced mother and her two children were original sample persons (rotation group 1), each with a first wave weight of 4,000.

The six time period, scenario combinations will be denoted by A.1, A.2, B.1, B.2, C.1 and C.2.

Note: We chose to determine the weights only for the longitudinal household that continues through the entire 1984-1985 period, which is marked with an asterisk above it. The other longitudinal households can also be weighted with all these procedures, except CM which applies only to the restricted universe.

Below is a schematic diagram of the example



A = divorced mother

b = c = divorced mother's child

D = divorced mother's widowed mother

E = widowed mother's new husband

Let W_{C1} = cross-sectional weight under scenario 1
W_{C2} = cross-sectional weight under scenario 2

Procedure BH

A.1., B.1., C.1. = W_{C1} for Abc = 0

A.2., B.2., C.2. = W_{C2} for Abc = 4,000

Procedure BI

A.1., C.1. = W_{C1} for AbcD = 0

B.1. = W_{C1} for AbcDE = (1/5) x 8,000 = 1,600

A.2., C.2. = W_{C2} for AbcD = (3/4) x 4,000 = 3,000

B.2. = W_{C2} for AbcDE = (3/5) x 4,000 + (1/5) x 8,000 = 4,000

Procedure CM

A.1. = W_{C1} for AbcD (the continuous members for the time period) = 0

B.1. = W_{C1} for AbDE (the continuous members for the time period)

= (1/4) x 8,000 = 8,000 = 2,000

C.1. = W_{C1} for AbD (the continuous members for the time period) = 0

A.2. = W_{C2} for AbcD (the continuous members for the time period)

(3/4) x 4,000 = 3,000

B.2. = W_{C2} for AbDE (the continuous members for the time period)

= (2/4) x 4,000 + (1/4) x 8,000 = 4,000

C.2. = W_{C2} for AbD (the continuous members for the time period)

= (2/3) x 4,000 = 2,666.67

Procedure AW

A.1. = [[(W_{C1} for AbcD) • 7 months] + [(W_{C1} for AbcDE) • 5 months]]/12 months

= [[(0) • 7] + [(1,600) • 5]]/12 = 666.67

B.1. = [[(W_{C1} for AbcDE) • 3 months] + [(W_{C1} for AbDE) • 9 months]]/12 months

= [[(1,600) • 3] + [(2,000) • 9]]/12 = 1,900

C.1. = [[(W_{C1} for AbcD) · 7 months] + [(W_{C1} for AbcDE) · 8 months] + [(W_{C1} for AbDE) · 9 months]]/24 months

= [[(0) · 7] + [(1,600) · 8] + [(2,000) · 9]]/24 = 1,283.33

A.2. = [[(W_{C2} for AbcD) · 7 months] + [(W_{C2} for AbcDE) · 5 months]]/12 months

= [[(3,000) · 7] + [(4,000) · 5]]/12 = 3,416.67

B.2. = [[(W_{C2} for AbcDE) · 3 months] + [(W_{C2} for AbDE) · 9 months]]/12 months

= [[(4,000) · 3] + [(4,000) · 9]]/12 = 4,000

C.2. = [[(W_{C2} for AbcD) · 7 months] + [(W_{C2} for AbcDE) · 8 months] + [(W_{C2} for AbDE) · 9 months]]/24 months

= [[(3,000) · 7] + [(4,000) · 8] + [(4,000) · 9]]/24 = 3,708.33

Procedure HW

A.1., B.1., C.1. = first wave cross-sectional weight for A = 0

A.2., B.2., C.2. = first wave cross-sectional weight for A = 4,000

5. POTENTIAL ADVANTAGES AND DISADVANTAGES

The ideal unbiased weighting procedure would provide a single set of weights applicable to any time interval, require no more data than were collected, and possess the minimum variance among all unbiased procedures. Unfortunately, no such procedure exists. The procedures described in Section 3 all fail one or more of these three criteria to various degrees. In this section, we explain the nature of the failures without explicitly comparing the procedures. That is done in Section 5.

Multiplicity of Weights. Some procedures have the advantage of assigning to each household a single weight which depends only on conditions as of the first reference month for the household and which is valid for every interval

that the household is in the universe. Other procedures have the disadvantage of sometimes producing different weights for the same household for different time intervals. (Procedures with this disadvantage could be modified so that only a single weight applies to any time interval, by computing for each household the weight appropriate for that procedure for the unrestricted universe and the 2 1/2 year time interval corresponding to the life of the panel. The weight obtained would also be used for any smaller subinterval for which the household is in the universe. However, weights obtained in this manner might not be able to be determined until the end of the life of the panel. This would make them difficult to use because we would have to wait until the last data from the panel were processed before estimates could be produced for any earlier time period. In any case, such weights would often lead to higher variances for short time intervals than weights developed specifically for the short time intervals.)

Unavailable Data Requirements. Most definition and procedure combinations require data from some households for time periods when the household is in existence but not in sample, that is for time periods for which interviews are not conducted for the household because no original sample people are members of the household. This needed data could be information for determining proper longitudinal weights or subject-matter information for use in tabulating the estimates. Some of this information is not collected for the 1984 panel of SIPP because of the current operational procedures. This is a consequence of the fact that agreement has not been reached on the longitudinal household definition to be used in SIPP. In this vacuum, operational procedures were determined mainly by considerations of difficulty and cost. Once a definition has been agreed on, depending on the nature of the unavailable data, it might be possible to change operational procedures for future SIPP panels so that

the required data are collected. To understand the problem with current operational procedures, consider the following situation. A household is longitudinal from month t_B to t_E. Original sample people are part of the longitudinal household only from month t_1 to t_2. If $t_B<t_1$, then some prior information may be unavailable. Revised operational procedures to obtain this information might involve retrospective questions, longer reference periods or proxy data on anyone who left the household before the first interview. If $t_2<t_E$, then some posterior information may be unavailable. Revised operational procedures might involve interviewing the household through t_E.

One of the important discriminants between the weighting procedures is how successfully they avoid the need for data from the period that the longitudinal household exists but is not in sample. (The need for such data is avoided by assigning zero weights to these problem households.) In terms of information needed for weighting, some procedures require only enough data to determine whether $t_B<t_1$, while others need to know t_B even when it is less than t_1. Similarly, some procedures only require knowledge of whether $t_2<t_E$, while others need to know t_E even when it is greater than t_2. Furthermore, besides this need for information for determination of weights, if any parameters other than the number of longitudinal households are to be estimated, then required subject-matter data may be missing as well, either before t_1, after t_2, or both.

While the problem of missing information is a serious one, it is not fatal. Procedures can be developed to compensate for the unavailable data. Specifically, the data collected on these households while they were in sample should be sufficient for performing imputation for existence/nonexistence outside the in-sample period and formation and/or dissolution dates. The imputed values can then be used to calculate weights for these households. These households can then be treated as noninterviews so that

the weights of mover households with similar demograpahic characteristics but with complete data receive increased weights while the deficient households themselves receives zero weights.

If the models underlying the procedures developed for adjusting for the missing information are true then it is still possible to obtain unbiased estimators, although now in a model-based sense. Furthermore, since the missing information that we are concerned with here is not caused by refusal to respond, modeling in this context might not suffer from the usually imperfect assumptions on similarity between respondents and nonrespondents that underlie any adjustments that use data from respondents to account for data missing from refusals. In addition, because of the longitudinal nature of the survey, there is generally a large amount of data available from the problem households that could be used in such adjustments. However, if the models are not perfect, then in general, the larger the proportion of data required that is unavailable, the greater the potential for serious bias problems.

Variances. In general, estimation procedures with the smallest variances are those that utilize available data intensively and tailor the weights to the specific time interval of interest. Unfortunately, as shall be seen in the next section, such procedures are often characterized by heavy needs for unavailable data which, as noted above, may impact unfavorably upon bias. Thus, there often is a direct trade-off between variance and the risk of bias. It will be difficult to weigh these factors against each other, since it appears that no single procedure will provide the correct balance for all of the multitude of characteristics that will be estimated by SIPP.

For use in the next section, we will define some labels for the advantages and disadvantages identified in the foregoing discussion. Let:

T_1 mean that a single longitudinal weight exists for each household, valid for all time intervals for which the household is in the universe, and which depends only on conditions which could be determined during the first interview,

T_2 mean the negation of T_1,

BW_1 mean that no data from the period preceeding the first interview are unavailable but required for weighting,

BW_2 mean that we need to know for weighting whether the longitudinal household existed before the first interview,

BW_3 mean that we need to know for weighting the conception date of the household (within the time interval of interest),

BD_1 mean that no subject-matter data from the period preceeding the first interview are unavailable but required,

BD_2 mean the negation of BD_1,

FW_1 mean that no data from the period following the last interview are unavailable but required for weighting,

FW_2 mean that we need to know for weighting the dissolution date of the household (within the time interval of interest),

FD_1 mean that no subject-matter data from the period following the last interview are unavailable but required,

FD_2 mean the negation of FD_1.

Note that T_1, BW_1, BD_1, FW_1 and FD_1 are the desirable properties.

6. DETAILED COMPARISONS OF ADVANTAGES AND DISADVANTAGES

Table 1 below presents advantages and disadvantages of each definition procedure and universe combination. A comparison of these features follows the table. Next, an explanation of each entry in the table is given. Finally, a discussion of data utilization, which is not in Table 1, is presented.

Table 1.

Features

Definition	Procedures	Universe	T_1	T_2	BW_1	BW_2	BW_3	BD_1	BD_2	FW_1	FW_2	FD_1	FD_2
No Change (NC)	All	Both	X		X			X		X		X	
Same Householder (SH)	Householder Weight (HW)	Both	X		X			X		X		X	
Same Householder (SH) Reciprocal Majority (RM) Shared Experiences (SE)	Beginning Date of Household (BH)	Unrestricted	X			X		X		X			X
SH, RM, SE	BH	Restricted	X			X		X			X		X
SH, RM, SE	Beginning Date of Time Interval (BI)	Unrestricted		X		X		X		X			X
SH, RM, SE	BI	Restricted		X	X			X			X		X
SH, RM, SE	Continuous Household Members (CM)	Restricted		X	X			X		X		X	
SH, RM, SE	Average Cross-Sectional Weight (AW)	Both		X			X		X		X		X

Comparison of Features in Table 1. As noted at the end of Section 4, T_1, BW_1, BD_1, FW_1, and FD_1 are the desirable properties. For the NC definition all five procedures considered here possess all these desirable properties, as does the HW procedure for the SH definition.

However, for the SH, RM, and SE definitions, and most other definitions too, the BH, BI, and CM procedures have different subsets of the set of desirable features, so that the procedure to be adopted depends, at least in part on the features deemed most important. AW possesses none of these desirable features for there three definitions. Its principal advantage lies in possible reductions in variances because of complete utilization of available data, which will be discussed later. BH has advantages T_1, BD_1, and FW_1 for the unrestricted universe, and T_1 and BD_1 for the restricted universe. The main reason for consideration of this procedure would be that it is the only one among BH, BI and CM that always has advantage T_1. BI has advantages BD_1 and FW_1 for the unrestricted universe and BW_1 and BD_1 for the restricted universe. Its principal advantage over BH is that for the restricted universe no retrospective questions need be asked. CM (which is only applicable to the restricted universe) possesses all desirable features except T_1, that is no information not currently collected is needed for this procedure. Recall, however, that CM had the disadvantage of being slightly biased as explained in Section 3.

Explanation of Entries in Table 1. All explanations presented below apply to both universes unless otherwise stated.

NC Definition, All Procedures. Since the composition of a household is unchanged throughout its period of existence under NC, we have the following two possibilities:

(a) No original sample people were in the household at any time during its period of existence, in which case the longitudinal household weight is zero for any time interval and procedure.

(b) One or more original sample people were in the household throughout its existence, in which case the beginning and ending dates of the household are known, as is the composition of the household and complete data for each month of its existence. Consequently, features BW_1, BD_1, FW_1, and FD_1 apply.

Furthermore, T_1 applies since procedures BH, BI, CM, and AW all reduce to the cross-sectional household weight at the beginning date of the household, while HW is the weight of the householder at the beginning date.

SH Definition, HW Procedure. The explanation is similar to the one given above, except now the two cases are: (a) The householder was not an original sample person. (b) The householder was an original sample person.

SH, RM, and SE Definitions, BH Procedure. T_1 is applicable, since by definition the weight is the cross-sectional household weight as of the beginning date of the household. BW_2 applies because the longitudinal household weight is the cross-sectional household weight as of the first month in sample if the household began that month, while otherwise the weight will be zero since there were no original sample people in the household when it began. (For the restricted universe, households which entered sample after the beginning of the time interval always receive a zero weight.)

BD_1 holds since all households with positive weights were in sample at their beginning date and no retrospective subject-matter data is therefore needed.

FW_1 holds for the unrestricted universe since the weight is determined at the beginning date of the household. However, for the restricted universe, it is necessary to know if the household continued to exist throughout the entire time interval because it receives a zero weight for the time interval if it did not continue. Under current procedures a household which no longer has any original sample person is not followed, and it would therefore generally

not be possible to determine if it remained in existence for the entire time interval. Consequently, FW_2 applies.

FD_2 applies since there would be missing data for all households with positive weights which continued to exist after there were no longer any original sample people present, which could happen for any of there three definitions.

SH, RM, and SE Definitions, BI Procedure. T_2 is applicable since time intervals with different beginning dates may yield different longitudinal weights. BW_1 applies for the restricted universe, since the longitudinal weight is the cross-sectional household weight as of the first month of the time interval for all households in sample that month, and zero for all other households. However, BW_2 applies for the unrestricted universe since longitudinal households that entered sample after the beginning of the time interval are treated as in the BH procedure.

BD_1 holds since any household with a positive weight was either in sample the first month of the time interval or the month that the household began, and consequently, no retrospective data are needed.

As in the BH procedure, and for the same reasons, FW_1 applies for the unrestricted universe, FW_2 for the restricted universe and FD_2 for both universes.

SH, RM, and SE Definitions, CM Procedure, Restricted Universe. T_2 is applicable since any two intervals may yield different longitudinal weights.

Furthermore, BW_1, BD_1, FW_1, and FD_1 apply. The explanation is similar to that given for the NC definition except now the two cases are:
(a) No original sample people were household members for the entire time interval. (b) At least one original sample person was a household member for the entire time interval.

SH, RM, and SE Definitions, AW Procedure. T_2 is applicable since any two time intervals may yield different longitudinal weights.

Any household that contained an original sample person for at least one month of the time interval receives a positive longitudinal weight for the unrestricted universe, while for the restricted universe it receives a positive weight if it also existed for the entire time interval. However, for either universe such a household might have existed for months when there were no original sample persons in the household, both before and after it came into sample. Hence BD_2 and FD_2 apply. Furthermore, in order to compute the longitudinal household weight it is necessary to determine if the household was in existence at the beginning and the end of the time interval for both universes, and in addition for the unrestricted universe, the beginning and ending dates if they are within the time interval. Hence BW_3 and FW_2 hold.

Utilization of Data. Having compared the procedures with respect to needs for unavailable data and the multiplicity of weights, we now turn our attention to variance. To compare the variance characteristics of the procedures we will focus on the amount of collected data that is used in obtaining estimates, since this is a primary determinant of variance. This discussion will also better illustrate the proportion of data needed for estimation that is unavailable for each procedure. In general, the greater this proportion is, the larger the burden is on any missing data procedure employed, with a resulting greater potential for bias problems. To make the comparison we show in Table 2, all 24 possible cases of how the data on a longitudinal household may be complete, partly available, or nonexistent for a particular time interval.

The symbols t_B, t_1, t_2, and t_E denote beginning date of household, first sample month, last sample month, and ending date of household respectively. The columns indicate different time intervals. Interval B is the interval of interest. Interval A is from t_B until the beginning of interval B, while interval C is from the end of interval B until t_E. The fifth case, for

example, is of a household that formed before interval B about which we are missing some data pertinent to the early part of interval B. The first nine cases comprise the restricted universe. The last 15 cases fill out the unrestricted universe. Each case is marked as having complete data, partial data, or no data. Of course, all of this is assuming perfect response. The only type of missingness that we are discussing here is that caused by operational procedures. On the right there is a column for each procedure with an "A" entered if it always uses the case, an "S" if it sometimes uses the case but not always (which will be explained in the discussion that follows), and a blank otherwise. These comparisons do not apply to the NC definition, for which all five procedures use all the complete cases and no other cases.

Table 2.

Data Utilization

	Interval A	Interval B	Interval C	Completeness	Procedure BH	BI	CM	AW	HW
1	$t_B=t_1$		$t_2<t_E$	perfect	A	A	S	A	S
2	$t_B<t_1$		$t_2<t_E$	perfect		A	S	A	
3	$t_B=t_1$	t_2	t_E	some missing	A	A		A	
4	$t_B<t_1$	t_2	t_E	some missing		A		A	
5	t_B	t_1	$t_2<t_E$	some missing				A	
6	t_B	t_1 t_2	t_E	some missing				A	
7	$t_B=t_1$ t_2		t_E	all missing	A				
8	$t_B<t_1$ t_2		t_E	all missing					
9	t_B		t_1 $t_2<t_E$	all missing					
10	$t_B=t_1$	$t_2=t_E$		perfect	A	A		A	S
11	$t_B<t_1$	$t_2=t_E$		perfect		A		A	
12		$t_B=t_1$	$t_2<t_E$	perfect	A	A		A	S
13		$t_B=t_1$ $t_2=t_E$		perfect	A	A		A	S
14	$t_B=t_1$	$t_2<t_E$		some missing	A	A		A	
15	$t_B<t_1$	$t_2<t_E$		some missing		A		A	
16		$t_B=t_1$ $t_2<t_E$		some missing	A	A		A	
17		$t_B=t_1$ t_2	t_E	some missing	A	A		A	
18	t_B	t_1 $t_2<t_E$		some missing				A	
19		$t_B<t_1$ $t_2<t_E$		some missing				A	
20		$t_B<t_1$ t_2	t_E	some missing				A	
21		$t_B<t_1$	$t_2<t_E$	some missing				A	
22	$t_B=t_1$ t_2	t_E		all missing	A				
23	$t_B<t_1$ t_2	t_E		all missing					
24		t_B	t_1 $t_2<t_E$	all missing					

The BH procedure uses the complete cases 1, 10, 12, and 13, but does not use the complete cases 2 and 11. It also uses the partial cases 3, 14, 16, and 17, and cases 7 and 22 for which there is no data in interval B. The BI procedure uses all the complete cases, more of the partial cases and none of the cases with no data. We thus think the BI procedure will tend to produce smaller variances than the BH procedure since it uses more of the available data. However, it is not clear in general which of these two procedures has the smaller proportion of needed data that is missing.

The CM procedure is appealing for the restricted universe since it uses all the complete cases (except in the rare situation when there is at least one original sample person present for every month of interval B, but none of them are present for the entire interval), and none of the other cases. It should thus have fairly small variances and has only the slight bias indicated in Section 3. However, it is not applicable to the unrestricted universe.

The HW procedure uses the same complete cases as the BH procedure, except it does not use these cases when the householder is not an original sample person, and it uses none of the other cases. However, it is not applicable to the RM, SE, and most other longitudinal household definitions.

The AW procedure is the most aggressive in utilizing partial data. It uses all the complete and partial cases while avoiding the cases with no data. Also note that it assigns smaller weights, in general, to the partial cases than the complete cases. We believe it will tend to produce the smallest variances for most definitions, particularly in the unrestricted universe, but also tends to have the highest proportion of data that is needed for estimation but unavailable.

7. ADJUSTMENTS OF ESTIMATES

In this section we will present some general ideas on adjustments to be made to the unbiased longitudinal household weights that would be obtained using any of the procedures described in Section 3. These should be considered only as preliminary thoughts, as many details remain to be worked out, and even the general approach is subject to change. The proposed procedures are somewhat analogous to the procedures used for cross-sectional estimates, and contain the following four components: an adjustment for the purpose of reducing between PSU sampling variability; an adjustment for household non-interview in second and subsequent waves; and a final adjustment to CPS estimates of the number of households by age-race-sex category of householder.

The first suggested step in the process of adjusting the unbiased weights does not actually begin with these weights, but instead alters the output of Section 3, so the resulting weights contain adjustments for first wave noninterview, and to reduce between PSU sampling variability. To do this, we simply alter the description in Section 3 of the first wave cross-sectional weight to now include these two adjustment factors in addition to the reciprocal of the probability of selection.

Two further adjustments would be performed on the weights resulting from the modification described in the previous paragraph. The need for the first adjustment would arise because there would be longitudinal households resulting from wave one respondent households for which there were missing data, not "completed" by imputation, for at least part of the time interval for which estimates are desired. This adjustment would redistribute the weights of such households to all households in the same weighting cells with complete data, in proportion to the weights of the households with complete data. In performing this adjustment it should be noted that in the case of

households for which complete contact is lost after some point, subsequent household compositional changes may alter the weights of the noninterview households, so it is not always clear what are the correct weights to redistribute. Imputation of these weights would appear to be necessary.

The final proposed adjustment would adjust the SIPP sample estimate of number of longitudinal households whose householder is in a given age-race-sex category to the CPS estimate. This would be accomplished by multiplying each household weight in the given cell by the ratio of the CPS estimate of the number of households in the cell to the SIPP estimate. (Family estimates could be controlled to CPS estimates by further dividing each cell into family and non-family household subcells. Even finer subdivision is also possible.) There are several possible approaches to computing this adjustment factor for each cell. The simplest would be to compute the factors at one month during the time interval in question, where the denominator of the ratio would be the sum of the weights of all longitudinal households in the cell in existence during that month, and then applying that same factor also to all other longitudinal households in the cell. (This was done in NMCUES (Whitmore, Cox, and Folsom 1982).) If this approach is taken then, in general, the SIPP and CPS estimates of the number of households in a given cell, and even the estimated total number of households in the universe, would not agree for any other month.

If it is required that the SIPP longitudinal household estimates in each cell agree with CPS estimates for every month in a time interval, then this could be accomplished by grouping the longitudinal households in each cell according to their pair of beginning and ending dates, and applying a different weighting factor for each such group. The values for these factors could be determined by considering them as variables in a mathematical

programming problem. This is described in detail by Judkins et al. (1984). Caution should be taken before adopting such a technique to control household weights for every month in a time interval. In certain situations no solution would be possible unless some weighting factors were allowed to be very large, or even negative. It may sometimes even occur that there is no solution even when there are no constraints on the weighting factors. Furthermore, slight changes in the objective function or the constraints might dramatically change some weighting factors. Finally, under some of the proposed definitions the householder in a longitudinal household may change, placing the household in a different age-race-sex cell, and requiring a modification of the procedure to account for this problem.

Some necessary imperfections in the CPS household control totals should also be noted. Although the CPS estimates of total individuals in a given age-race-sex category are themselves controlled to independent demographic estimates which have no sampling variability, the number of householders in each category is not controlled in this manner. This is troubling because the process which yields the CPS estimates of households is subject to unknown biases. Despite this, it is felt that this use of CPS estimates in adjusting SIPP data would reduce total sampling variability and many biases because of the combination of the demographic estimate controls and the larger size of the CPS sample.

REFERENCES

Dicker, Marvin and Casday, Robert J. (1982), "A Reciprocal Rule Model for a Longitudinal Family Unit," American Statistical Association - Proceedings of the Social Statistics Association.

Judkins, D.R., Hubble, D.L., Dorsch, J.A., McMillen, D.B., and Ernst, L.R. (1984), "Weighting of Persons for SIPP Longitudinal Tabulations," American Statistical Association - Proceedings of the Section on Survey Research Methods, to appear.

Kasprzyk, Daniel and Kalton, Graham (1983), "Longitudinal Weighting in the ISDP," Technical Conceptual and Administrative Lessons of the ISDP, Social Science Research Council, New York.

McMillen, David B. and Herriot, Roger A. (1984), "Toward a Longitudinal Definition of Households," SIPP Working Paper Series, No. 402, Bureau of the Census, Washington, D.C.

Nelson, D.D., McMillen, D.B., and Kasprzyk, D. (1984),"An Overview of the Survey of Income and Program Participation", SIPP Working Paper Series, No. 401, Bureau of the Census, Washington, D.C.

Whitmore, R.W., Cox, B.G., and Folsom R.E. (1982), "Family Unit Weighting Methodology for the National Household Survey Component of the National Medical Care Utilization and Expenditure Survey," Research Triangle Institute report.

EARLY INDICATIONS OF ITEM NONRESPONSE ON THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

John F. Coder and Angela M. Feldman
U.S. Bureau of the Census

For Presentation at the
American Statistical Association
Annual Meetings - August 1984

EARLY INDICATIONS OF ITEM NONRESPONSE ON THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

by John F. Coder and Angela M. Feldman

Introduction

The Survey of Income and Program Participation (SIPP) promises to become the most important source of data for measuring the level of and changes in the economic well-being of the U.S. population. Collection of these data began in the fall of 1983. The survey design for the initial sample of 25,900 housing units in the noninstitutional population, calls for each household to be interviewed at 4-month intervals over a 2-1/2 year period. The sample is divided into 4 rotations or panels of equal size and one panel is interviewed in each month throughout this period resulting in a total of eight personal contacts by Census interviewers for each sample household.

The first interviews in this new survey were conducted during October, November, and December of 1983, and January 1984. The questionnaire used to collect information in the initial interview concentrates on labor force participation and sources and amounts of income. Most data is recorded separately by month for the 4-month reference period ending in the month prior to the month of interview. For example, data collected in the October 1983 interviews covered the June through September period. Most interviews were completed during the first 2 weeks of the interview month.

The primary purpose of this paper is to present some preliminary indications of the item nonresponse rates for the first interviews of SIPP. These rates of nonresponse cover labor force, income recipiency, and income amounts. The effect of self or proxy respondents on nonresponse rates is discussed for

a selected group of items. Some data on other aspects of the survey have also been included. These are overall household noninterview rates, average times required for interviews, and use of callback procedures to obtain missing information.

Item Nonresponse

Item nonresponse is defined in this paper to mean a missing answer to a specific question that should have been answered. Item nonresponse can result for many reasons, the most frequent being lack of knowledge by the respondent, i.e., "Don't Knows," and refusals to answer. Nonresponse can also result when the interviewer fails to record a response in the correct location or follows an incorrect path within the questionnaire design.

Labor Force Items--Table 1 shows preliminary nonresponse rates for items 2a, 2b, 4, 5a, 5b, 6a, 6b, 6c, 7a, 7b, and 8a of the labor force and recipiency section on the first interview questionnaire. The questions themselves are shown in Figure 1.

In general, the nonresponse rates for the labor force questions were low (see table 1). The nonresponse rate on item 2a, incidence of looking for work or on layoff for persons who did not work at all during the reference period (nonworkers) was only 0.4 percent. About 6.7 percent of the nonworkers reporting looking or on layoff had a nonresponse for item 2b, the number of weeks spent looking or on layoff. The comparable nonresponse rates for workers were 1.0 percent for incidence of looking or on layoff (item 7a) and 3.2 percent for item 7b, the number of weeks spent looking or on layoff. The nonresponse rate for item 4, asking if the respondent held a job or business during the entire 4-month reference period, was less than 0.1 percent.

One of the questions with a relatively high nonresponse rate in the labor force section was item 5b covering the number of weeks absent without pay for persons having a job for the entire period. The nonresponse rate for this question was 11.6 percent.

Item 8a is the question covering the number of hours usually worked per week during the 4-month period. This critical data item was missing for 1.3 percent of the 25,510 sample persons reporting a job or business during the reference period.

Income Recipiency.--A major portion of the questionnaire was designed to determine the sources of income received during the 4-month period by each household member age 15 years old and over. A total of 52 different income sources (other than earnings from employment) were covered in the survey. Tables 2 and 3 show income recipiency nonresponse rates and ratios of nonresponses to "YES" responses for SIPP and the March 1983 CPS for a selected group of income types. The rates refer to the 4-month reference period for SIPP and calendar year 1982 for the March CPS.

The nonresponse rates for SIPP are extremely low and vary only slightly by rotation. The nonresponse rate on recipiency for SIPP ranged from less than 0.1 for Aid to Families with Dependent Children and private pensions to 1.3 percent for stocks or mutual funds. In contrast, the rates for the March 1983 CPS clustered around the 10-percent level. These rates for the March CPS are largely attributable to the 7 percent household noninterview rate on the income supplement questionnaire.

The last two columns of table 3 show the ratios of nonresponses to "YES" responses for SIPP and the March CPS. This measure of nonresponse may be better

than the overall nonresponse rate because it provides a measure that is relative to the size of the recipient universe. The March CPS ratios are again much higher than those encountered in the first interview of SIPP. This difference is also related to the 7 percent March supplement noninterview rate. Given this fixed nonresponse rate the ratio is inversely related to the proportion of the population receiving a specific income type. This is evident by the large ratio of 4.01 for Aid to Families with Dependent Children. The ratio itself means that, in this case, the number of nonresponses and, therefore, imputations required exceeded the number of "YES's" by a factor of 4 to 1.

Hourly Wage Rates.--The nonresponse rates on hourly wages are shown in table 4. These rates are shown separately by type of respondent. The nonresponse rate was 9.5 percent overall, 5.1 percent for self response and 16.7 percent for proxy response. The overall nonresponse rate for hourly wages increased from the 7.8 percent level in October to 10.5 percent in January. This resulted mainly from an increase in the nonresponse rate for proxy responses of from 13.8 percent in October to 19.2 percent in January. Approximately 62 percent of the respondents were "self."

Monthly Wage or Salary Income.--Table 5 contains the nonresponse rates for the monthly amounts of wage and salary income. The nonresponse rate overall averaged about 6.2 percent for the initial SIPP interviews. The rate for self respondents, which accounted for 64 percent of the total, was lower, 4.6 percent, while the rate for proxy respondents was 9.0 percent. The 9.0-percent nonresponse rate for proxy interviews on monthly earnings amounts was considerably lower than the comparable rate of 16.7 percent for hourly wage amounts.

Nonresponse rates increased from 5.4 percent to 6.7 percent between October and January.

Self-Employment Income.--Nonresponse rates for self-employment income have traditionally exceeded those for most income types. The items in the self-employment section of the SIPP questionnaire cover monthly amounts of "salary" and other income received by owners of businesses, professional practices, farms, etc. The question is not designed to obtain estimates of the business's net profit on a monthly accounting period. An additional question was included covering estimated net profit for the entire 4-month reference period. The nonresponse rate overall for the monthly salary or other income received by the self-employed was 14.0 percent (see table 6). The nonresponse rate for proxy interviews exceeded that of self-responses by a considerable margin. The rate for proxy interviews was 22.3 percent compared to 9.8 percent for self responses. Nonresponse rates were slightly higher in January than October, increasing from 13.6 percent to 15.1 percent. About two-thirds of respondents for this item were "self."

Interest Income.--Table 7 contains nonresponse rates for interest amounts received during the SIPP 4-month reference period. These rates cover the interest amount received from one or more of the following sources: 1) regular or passbook savings, 2) money market deposit accounts, 3) certificates of deposit, or other savings certificates, and 4) NOW accounts or other interest earning checking accounts. The nonresponse rate for interest income from these sources was 34.6 percent. The rate in January was 35.4 percent, somewhat higher than the 32.6 percent for October. About 4 percent of the total number of nonresponses on interest amounts can be attributed to refusals. The remainder

were mainly categorized as "Don't Knows." A "Don't Know" response to interest income was followed by a question to obtain the balance or amount in the account. The nonresponse rates for this item are also shown in table 7. The nonresponse rate for balances in savings was 24.2 percent. In combination these two nonresponse rates indicate that both the interest amount and the balance amount were missing in only about 13.3 percent of the sample cases for these sources of interest income.

Dividend Income.--The questions covering the amount of dividend income received were divided into two categories, those dividends actually received and those credited against a margin account or automatically reinvested in additional shares of stock. As indicated by the data in table 8, the nonresponse rates for these two categories differ significantly. The rate for dividends actually received was 9.4 percent. The rate for dividends credited was 30.7 percent.

Noninterview Rates

The noninterview rate is a measure of the proportion of occupied housing units, i.e., those eligible for interview, for which interviews were not obtained. As mentioned earlier the total sample size for the 1983 SIPP was about 25,900 housing units. Of this total about 4,600 were not eligible for interview. These ineligible units were found to be vacant, demolished, under construction, or unoccupied for other reasons. This left 19,900 households eligible to be contacted. Interviews were not obtained for 4.8 percent of this group (see table 9). Most noninterviews, about 77 percent, were refusals to participate. The remainder of the total noninterview rate consisted of

situations classified as "no one home" and "temporarily absent." These classifications were assigned after repeated visits failed to yield a contact.

The noninterview rate varied considerably by region of the Country. The lowest noninterview rate was 2.4 percent from the Kansas City Regional Office that covers Kansas, Missouri, Iowa, Minnesota, and Wisconsin. The highest noninterview rate was 10.1 percent from the New York Regional Office covering the parts of New York and New Jersey in the vicinity of New York City.

There was slight variation in the noninterview rates by month of interview, however, there was no apparent trend. The rate for the first month of interview was 5.1 percent compared to 4.3 percent, 5.2 percent, and 4.8 percent in the succeeding 3 months, respectively. The overall noninterview rate of 4.8 percent was somewhat higher than the March 1983 CPS rate of 4.4 percent. The rate for SIPP was, however, lower than the 5.4 percent noninterview rate for the panel coming into the March 1983 CPS for the first time. As noted earlier, about 7.0 percent of the March CPS sample households completed the monthly labor force questions but were noninterviews on the income supplement. These cases are in addition to the 4.4 percent household noninterviews.

Callback Items

The design of the SIPP questionnaire incorporated procedures for following up on missing responses to items identified as either especially important to the overall quality of the survey data or with previously noted high nonresponse rates. The first step in this process was the determination that the answer to the designated question would be available from another household member not present at the time of the interview or at a later date. If so, the interviewers, in most cases, called back by telephone to obtain the missing information. The data in table 10 summarize use of the callback system.

The callback system appears to be most effective for obtaining missing data on amounts of monthly wage and salary income. About 600 cases were marked for callback for these amounts. The procedure obtained responses to the missing earnings amounts in about 7-out-of-10 cases.

Use of the callback was less successful in obtaining missing amounts for the other income sources. Slightly more than half (54 percent) of the callbacks were successful for obtaining data for the monthly amount of salary and other income received from self-employment. Attempts to follow up on amounts of interest and dividend income from various sources proved to be even less effective. About 45 percent of the respondents were able to supply an amount when contacted by an interviewer. Use of the callback procedures appears to have declined between the October and January interviews. The number of cases marked for follow-up in January were significantly lower than October for each income type. While less frequent use of the callback might have been related to a reduced need for follow-up, nonresponse rates for these income types tended to increase between October and Janaury, indicating the opposite.

Interview Time

The time required to conduct an initial SIPP interview is potentially quite long given the number of questions. Obviously households with a large number of adult members, those 15 years old and over, are those that are exposed to the longest overall interview times, on average. The data in table 11 provide the first estimates of interview times based directly on times entered on each person's questionniare by the interviewers. The time required to complete the household control card and roster was added to the interview time on the first questionnaire for the household. These estimates are shown by size of household for the first interview period of SIPP.

The median interview time was 43 minutes for all households in the first interview. The median interview time declined steadily from 48 minutes in October to 41 minutes in January. The median household interview time for 1-person households was about one-half hour while that for 4-person households was one hour and ten minutes. Households with 5, 6, and 7 or more members required proportionally more time for interviews.

Summary

This examination of some of the early "returns" from the 1983 SIPP are, for the most part, encouraging. The household noninterview rate was lower than most had anticipated. The item nonresponse rates were much lower than those experienced in the March CPS. Proxy responses caused significantly higher nonresponse rates for some of the key items studied.

There is reason for concern, however, in several areas and these should be watched closely. The first is the general trend toward higher nonresponse rates between October and January interviews. The second is the relatively high noninterview rate for the New York area. While this is consistent with our experiences in other surveys, this rate should be monitored closely as will the rates in the other regions.

The next step in the evaluation of the 1983 SIPP data will be comparison of the survey estimates of income recipients with figures derived from program statistics and other independent sources. This analysis will provide a very important look at the magnitude of survey underreporting, a major concern of SIPP and other household income surveys.

Figure 1. Selected Labor Force Questions

NONWORKERS

2a. Even though ... did not have a job during this period, did ... spend any time looking for work or on layoff from a job?

|__| YES -- ASK 2b

|__| NO

2b. In which weeks was ... looking for work or on layoff from a job?

WORKERS

4. Did ... have a job or business, either full or part time, during EACH of the weeks in this period?

|__| YES -- ASK 5a

|__| NO -- ASK 6a

5a. Was ... absent without pay from ...'s job or business for any FULL weeks during the 4-month period?

|__| YES -- ASK 5b

|__| NO

5b. In which weeks was ... absent without pay?

WORKERS WITH WEEKS WITHOUT A JOB OR BUSINESS

6a. In which weeks did ... have a job or business?

6b. Was ... absent from work for any full weeks without pay?

|__| YES --ASK 6c

|__| NO

6c. In which weeks was ... absent without pay?

7a. During the weeks that ... did not have a job did ... spend any time looking for work or on layoff?

|__| YES -- ASK 7b

|__| NO

7b. In which of these weeks was ... looking for work or on layoff from a job?

WORKERS

8a. In the weeks that ... worked during the 4-month period, how many hours did ... usually work per week?

Table 1. Selected Item Nonresponse Rates for the Labor Force Items on the 1983 SIPP: Interview No. 1

Item	Total	Rotation			
		One	Two	Three	Four
2a	0.4	0.4	0.4	0.4	0.3
2b	6.7	8.2	6.8	5.9	5.9
4	0.1	0.1	0.1	(Z)	0.1
5a	0.1	0.1	0.1	0.1	0.1
5b	11.6	12.6	11.0	8.2	14.4
6a	2.2	2.9	2.0	1.9	1.8
6b	3.3	6.6	2.3	1.8	1.4
6c	6.8	2.1	12.2	3.3	10.5
7a	1.0	0.9	1.0	1.1	0.9
7b	3.2	4.7	3.7	2.0	2.0
8a	1.3	1.3	1.3	1.2	1.2

Z Less than .05 percent.

Table 2. Selected Item Nonresponse Rates for Income Recipiency During the 4-month Reference Period on the 1983 SIPP: Interview No. 1

Income type	Total	Rotation			
		One	Two	Three	Four
Social Security.....	0.6	0.6	0.6	0.5	0.5
Unemployment compensation..	0.1	0.1	0.1	0.1	0.1
Veteran's payments.....	0.2	0.2	0.2	0.2	0.2
Aid to Families with Dependent Children.....	(Z)	(Z)	(Z)	(Z)	(Z)
Food stamps.....	0.3	0.4	0.4	0.2	0.2
Private pensions.....	(Z)	(Z)	(Z)	0.1	(Z)
Savings accounts.....	1.0	0.8	0.8	1.1	1.1
Shares of stock or mutual funds.....	1.3	1.4	1.1	1.3	1.3
Rental property.....	1.0	1.0	0.7	0.9	1.1

Z - Less than .05 percent.

Table 3. Selected Income Nonresponse Rates from the March 1983 CPS, Ratio of Nonresponses to "YES" Responses for the March 1983 CPS, and Ratio of Nonresponses to "YES" Responses for Interview NO. 1 of the 1983 SIPP

Income type	March 1983 CPS nonresponse rate	March 1983 CPS ratio of nonresponses to "YES's"	1983 SIPP ratio of nonresponses to "YES's"
Social Security.....	9.6	0.61	.03
Unemployment compensation..	9.6	1.16	.03
Veteran's payments.....	9.6	1.14	.10
Aid to Families with Dependent Children.....	9.7	4.28	.01
Food stamps.....	6.4	0.84	.07
Private pensions.....	9.6	1.64	.01
Savings accounts.....	10.4	.215	.02
Shares of stock or mutual funds.....	9.7	0.69	.09
Rental property.....	9.7	0.66	.13

Table 4. Nonresponse Rates on Hourly Wage Rate by Type of Respondent for the 1983 SIPP: Interview No. 1

Type of respondent	Total	Rotation			
		One	Two	Three	Four
Total.....	9.5	7.8	9.3	10.4	10.5
Self.....	5.1	4.1	4.7	5.9	5.6
Proxy.....	16.7	13.8	16.1	18.0	19.2
Proportion of Self Responses.	.62	.62	.60	.63	.64

Table 5. Nonresponse Rates on Monthly Wage and Salary Income by Type of Respondent for the 1983 SIPP: Interview No. 1

Type of respondent	Total	Rotation			
		One	Two	Three	Four
Total.....	6.2	5.4	5.8	6.8	6.7
Self.....	4.6	4.2	4.3	4.9	4.9
Proxy.....	9.0	7.6	8.4	10.2	10.1
Proportion of Self Responses.	.64	.63	.63	.64	.65

Table 6. Nonresponse Rates on Monthly Amounts of Self-Employment Income for the 1983 SIPP: Interview No. 1

Type of respondent	Total	Rotation			
		One	Two	Three	Four
Total.....	14.0	13.6	12.6	14.6	15.1
Self.....	9.8	9.5	9.7	9.6	10.2
Proxy.....	22.3	21.4	18.6	24.3	24.7
Proportion of Self Responses.	.66	.65	.67	.66	.66

Table 7. Nonresponse Rates for Amounts of Interest Income from the 1983 SIPP: Interview No. 1

Item	Total	Rotation			
		One	Two	Three	Four
Interest amount.....	34.6	32.6	33.8	37.1	35.4
Percent refusals.....	4.2	4.1	4.0	4.6	4.1
Balance amount.....	24.2	23.6	24.1	24.9	24.1

Table 8. Nonresponse Rates for Amounts of Dividend Income for the 1983 SIPP: Interview No. 1

Item	Total	Rotation			
		One	Two	Three	Four
Dividends received.....	9.4	10.3	8.3	9.8	9.3
Dividends credited.....	30.7	28.2	33.8	30.1	30.5

Table 9. Household Noninterview Rates by Regional Office for the 1983 SIPP: Interview No.1

Item	Total	Rotation			
		One	Two	Three	Four
Total.....	4.8	5.1	4.3	5.2	4.8
Boston.....	3.8	2.9	2.5	5.4	4.6
New York.....	10.1	13.3	8.3	10.8	8.4
Philadelphia.....	3.0	2.0	3.4	2.5	4.1
Detroit.....	4.1	3.0	3.6	5.4	4.1
Chicago.....	4.8	5.0	3.4	5.7	5.0
Kansas City.....	2.4	1.6	1.6	4.0	2.5
Seattle.....	4.7	5.1	4.4	5.2	4.3
Charlotte.....	3.5	4.3	2.7	2.8	3.8
Atlanta.....	4.9	5.4	5.0	5.2	4.2
Dallas.....	5.1	5.0	5.1	4.6	5.8
Denver.....	5.3	6.1	5.7	4.1	5.5
Los Angeles.....	7.5	9.3	6.2	8.9	5.8

Table 10. Success Rates of Callback Items

Item	Total	Rotation			
		One	Two	Three	Four
Success Rates					
Wages and salary.....	71.0	76.2	76.9	70.0	59.0
Self-employment.....	54.0	58.6	55.0	48.3	54.5
Interest and dividends.	44.8	48.4	49.6	38.2	40.8
Number of Callbacks					
Wages and salary.....	599	172	143	150	134
Self-employment.....	100	29	20	29	22
Interest and dividends.	582	192	139	131	120

Table 11. Median Household Interview Times by Number of Members 15 Years Old and Over from the 1983 SIPP: Interiew No. 1

Number of persons	Total	Rotation			
		One	Two	Three	Four
Total.....	43	48	44	42	41
One.....	29	33	30	26	26
Two.....	44	50	45	42	41
Three.....	57	64	57	55	55
Four.....	70	76	72	67	66
Five.....	83	90	81	84	77
Six.....	98	105	111	101	71
Seven or more.....	113	114	(B)	120	94

B Less than 10 sample households.

Survey of Income and Program Participation

Working Paper Series

No. 8407

Programs Presented in the
Survey of Income and Program Participation
Session IV
at the Annual Meeting of the
American Statistical Association
in Philadelphia, PA
August 13-16, 1984
(No. 8407)

U.S. Department of Commerce
Malcolm Baldrige, Secretary
Clarence J. Brown, Deputy Secretary
Sidney Jones, Under Secretary for
Economic Affairs

BUREAU OF THE CENSUS
John G. Keane,
Director

U.S. BUREAU OF THE CENSUS
John G. Keane, Director
C.L. Kincannon, Deputy Director
William P. Butz, Associate Director
for Demographic Fields

POPULATION DIVISION
Roger A. Herriot, Chief

Acknowledgements

This publication is composed of papers prepared by many different authors for presentation at the American Statistical Association on August 13-16, 1984. We would like to thank these authors for their cooperation in making the papers available for publication. Clerical and editorial assistance was provided by Hazel Beaton, Mary Kisner, and Delma Frankel.

Suggested Citation

"Papers Presented at the Survey of Income and Program Participation session IV at the annual meeting of the American Statistical Association in Philadelphia, PA, August 13-16, 1984," SIPP Working Paper Series No. 8407. U.S. Bureau of the Census, Washington, D.C. 1984.

Preface

This report is comprised of five papers featured in the "Survey of Income and Program Participation" session IV, one of two in the Survey of Research Methods Section of the annual meeting of the American Statistical Association.

This session covered a range of topics, both methodological and substantive, about longitudinal surveys and the Survey of Income and Program Participation (SIPP).

SIPP is a new Census Bureau survey collecting data that will help measure income distribution and poverty throughout the country more accurately. These data will be used to study Federal and state aid programs (such as food stamps, welfare, Medicaid, and subsidized housing), to estimate future program costs and coverage, and to assess the effects of proposed changes in program eligibility rules or benefit levels.

Households in the survey will be interviewed at 4-month intervals over a period of 2 1/2 years. The reference period will be the 4 months preceding the interview. In all, about 20,000 households will be interviewed, approximately 5,000 each month. Field operations will be handled through our 12 regional offices.

Recurring questions will deal with employment, types of income, and noncash benefits. Periodic questions will be added dealing with school enrollment, marital history, migration, disability, and other topics. Special supplemental questions will also be added to the SIPP questionnaire.

These papers discuss SIPP and its predecessor, the Income Survey Development Program (ISDP), an experimental program designed to test procedures used in conducting SIPP.

Session: Survey of Income and Program Participation IV

Session Chair: Graham Kalton
Survey Research Center
University of Michigan

Papers:

"Month-to-Month Income Recipiency Changes in the ISDP."
Written by J. C. Moore and D. Kasprzyk, Bureau of the Census.

(Examines a tendency for reported program turnover in the 1979 Panel to occur between waves more often than within waves.)

"Findings from the Student Follow-Up Investigation of the 1979 ISDP."
Written by A. M. Roman and D. V. O'Brien, Bureau of the Census.

(Discusses objectives, design, and results of an investigation to ascertain whether parents are reliable proxies for their children who are students living away from home.)

"The ISDP 1979 Research Panel as a Methodological Survey: Implications for Substantive Analysis."
Written by R. A. Kulka, Research Triangle Institute.

(Discusses the potential for the data in this panel to address social, economic, and policy research issues either by cross-sectional or longitudinal analysis.)

"Some Data Collection Issues for Panel Surveys with Application to SIPP."
Written by A. Jean and E. K. McArthur, Bureau of the Census.

(Describes operational methods for dealing with household changes during the survey.)

"Managing the Data from the 1979 ISDP."
Written by P. Doyle and C. Citro, Mathematica Policy Research, Inc.

(Discusses data access and analysis problems in the ISDP and the use of a data base management system to resolve them.)

MONTH-TO-MONTH INCOME RECIPIENCY CHANGES IN THE ISDP

J.C. Moore
and
D. Kasprzyk

Bureau of the Census
Washington, D.C. 20233

For presentation at the annual meeting
of the American Statistical Association,
Philadelphia, Pa., August 13-16, 1984.

MONTH-TO-MONTH RECIPIENCY TURNOVER IN THE ISDP

Jeffrey C. Moore and Daniel Kasprzyk
U. S. Bureau of the Census

The major impetus to the development of the Survey of Income and Program Participation (SIPP) was the need for more detailed and better quality income data than were available through current survey programs--most notably, the March income supplement to the Current Population Survey (CPS) (David, 1983; Ycas and Lininger, 1981). The SIPP itself has only been in the field since October of 1983, so there are not yet sufficient data for a thorough assessment of its performance. However, the precursor to the SIPP, the Income Survey Development Program (ISDP), is an available and underutilized data source offering a wealth of information to researchers with interests in a wide range of SIPP-related issues.

Background

This paper uses the 1979 Panel of the ISDP to examine a particular data quality problem concerning month-to-month turnover in the receipt of various income types. The basic question, first raised by Czajka (1982), is as follows: given six monthly observations over two consecutive survey waves (each of which covers retrospectively a 3-month period), what is the pattern of recipiency turnover in the resulting five pairs of months? Czajka's interpretation of tables prepared for another purpose by Lepkowski and Kalton (1981) was that in survey waves 1 and 2 of the 1979 panel there was "a pronounced tendency for reported program turnover to occur between waves more often than within waves--i.e., between months three and four rather than the four other pairs of months" (p. 93). Moore (1983), however, in a quantitative analysis of the Lepkowski and Kalton tables, failed to find the effect suggested by Czajka.[1]

This discrepancy between the two investigations is attributable to differing interpretations of one of the response indicators in the tables--specifically, whether a particular code indicated "no data" (i.e., a case which could not be matched across the two waves) or "no receipt." Notwithstanding this confusion, two additional factors argued strongly for a more careful examination of the issue. First was the issue of completeness. For their work, Lepkowski and Kalton linked only the first two waves of the 1979 panel, leaving untouched waves 3, 4, and 5. A second shortcoming had to do with the quality of the linking operation itself. Lepkowski and Kalton had at their disposal only an early version of the ISDP data file, which contained numerous errors in the person identifier code crucial to the linking of survey records across waves.[2]

Subsequent work carried out by Mathematica Policy Research, Inc., apparently corrected the problems with the person identifiers, resulting in the creation of a linked data file which had substantially more matches than the earlier file produced by the Michigan group. In addition, all five relevant waves of the 1979 Panel were included in the linking operation. The remainder of this paper analyzes and discusses tabulations derived from the later "definitive" edition of the 1979 ISDP data file to address more conclusively the issue of within-wave versus between-wave month-to-month income recipiency turnover.

Real underlying trends: Since this investigation is without the benefit of external validating information, we cannot demonstrate conclusively that the observed results indicate "error" as opposed to reflecting accurately real underlying trends in the events being measured. Two facts, however, render the latter hypothesis untenable: 1) a change in economic conditions or eligibility rules could produce an increase in recipiency turnover at a particular point in time, but it is difficult to imagine this happening periodically for a wide range of income types over an extended period of time; 2) the staggered interviewing schedule for the 1979 ISDP Panel (see Ycas and Lininger, 1981) further reduces this likelihood, since each calendar month over the life of the panel served as the first reference month of a wave for one set of respondents, the second reference month for another set, and the third month for a third set. In other words, each reference month in a survey wave combines data from three calendar months, so that any real change effects are present only in diluted form in three reference months.

Edit and imputation procedures: Three processing procedures possibly contributed to greater recipiency turnover between waves than within waves: reformatting edits to simplify and make consistent various data fields, imputation for person nonresponse, and imputation for item nonresponse.

The only known problem with the reformatting edits is that they were carried out independently for each wave; incorrect resolutions in the name of consistency thus may have artificially reduced turnover within waves, while reporting inconsistencies between waves were ignored. Another edit decision which may have contributed to the phenomenon of less turnover within waves than between waves was the following: if at least one "yes" was reported for an income type, and/or if at least one monthly amount was a valid nonzero amount, then any blank monthly recipiency indicators were set to "yes" and any blank monthly amounts were imputed using the average of the amounts reported in other months. The obvious effect of such a procedure is to reduce the apparent amount of change within a wave. Unfortunately, these edits were not identified on the data file. As a result, the extent to which they affected the results presented here is not known, although their combined impact is likely to be small.

Another possible contributor to the observed effect is the treatment of person noninterviews within interviewed households. Because there were, in fact, few such cases (only 298 in Wave 1), an imputation procedure was developed to substitute complete person records for the otherwise missing data. The procedure used reported demographic data as matching variables in a hot-deck assignment. Since each wave's data were processed independently, it is highly unlikely that an individual who was a nonrespondent in each of two consecutive waves would receive the same imputation donor for both waves. Consequently, some spurious wave-to-wave change could occur solely as an artifact of the independent processing.

The same argument applies to the case of item nonresponse within a person's record. The presence of valid data in one wave and the absence of valid data in the next (or vice versa) suggests possible problems for between-wave analyses because the ISDP imputation system did not take previous (subsequent) reporting patterns into account. In addition, if a respondent did not provide information for a specific item on two successive waves of interviewing, it is likely that different imputation donors provided the missing data in each wave.

Mismatches: Technically, of course, although respondents do report month-to-month turnover within a survey wave, it is incorrect to refer to respondents' "reports" of between-wave turnover. These events are created by the computerized process which links together the data for specific individuals across survey waves. To the extent that people are incorrectly linked, a certain amount of artifactual turnover may appear in the month-pair which connects the two waves. Preliminary simulation work suggests that mismatching need not be extensive to produce within-wave versus between-wave differences of the magnitudes observed in Table 1. In fact, for most of the income types in this paper, a mismatch rate of 3 percent or less would produce an apparent increase in turnover quite comparable to the observed increase from within-wave month-pairs to between-wave pairs.

It is impossible after the fact to determine the impact of person mismatches on the estimates of between-wave turnover in the 1979 panel. Returning to the discrepancy between the early Lepkowski and Kalton data and the subsequent refined file, one intriguing possibility is that although the former produced fewer matches than the latter, the matches that were completed may have been relatively error-free. If this were the case--that is, if the Michigan group somehow skimmed off the definite matches--then the appearance of heightened between-wave turnover in the later data file may simply reflect increased match errors. Clearly, evaluating the impact of match errors in turnover estimates from the SIPP will require maintaining data on the quality of the match for each person, perhaps in the form of a scale showing the number of variables which were identical across the linked waves.

Response error: Perhaps the most common explanation for the effects observed in this paper involves some form of recall bias. This was certainly Czajka's (1982) assumption. Presumably, a gestalt-like process operates in response to imperfect recall, leading respondents to report receipt for the entire 3-month period of a single wave as having been more stable than it really was. Such a process would work in two ways to produce more reports of between-wave than within-wave turnover: first, by reducing the number of within-wave turnover episodes (see Example 1); and second, by shifting the occurrence of turnover episodes to the between-wave period (Example 2).

		wave n			wave n+1		
	month:	1	2	3	4	5	6
Example 1	actual receipt:	yes	no	yes	no	yes	no
	reported receipt:	yes	yes	yes	no	no	no
Example 2	actual receipt:	yes	yes	yes	yes	no	no
	reported receipt:	yes	yes	yes	no	no	no

Although it is impossible with the available data to evaluate these notions directly, other research has demonstrated effects which appear to be related to the processes hypothesized to be at work here. Goudreau, Oberheu, and Vaughan (1984) report two results of interest from a survey of known AFDC

recipients. First, those who failed to report receipt were likely to have received AFDC income for only part of the reference period of the survey. And second, the most common error in reporting income amounts was the tendency to report "the most recent payment for all three months of the reference period when payments actually varied" (p. 184).

A second, related response error possibility can be examined using the present data. According to this explanation, misreports of the type described above, while perhaps representing a general human tendency, are even more likely to occur when the respondent and the subject of the report are not the same person, and especially when different respondents provide the data for two consecutive survey waves. Table 2 summarizes the data regarding the role of proxy response in general, and changing respondents specifically, on elevated between-wave turnover. The results do not present a simple picture, but there is no evidence that self-response in consecutive waves erases the general effect observed in this paper. Note that with only one exception, all differences in column (c) are positive; that is, between-wave turnover is consistently greater than within-wave turnover even when attention is restricted to the constant self-response group.

Nor, in fact, is there consistent support for the weaker argument that self-response might at least reduce between-wave/within-wave turnover discrepancies. As shown in columns (j) and (m), the weight of the evidence is in the opposite direction. Only for the two earned income categories does proxy involvement strongly and consistently produce greater differences as compared to constant self-response.

Why the two general income types produce such disparate results is not clear. A plausible partial explanation--at least for the both-self/mixed-self-and-proxy comparison--is that a true change in recipiency for earned income also changes a person's availability for interview. For example, those who are not employed may be more readily available to be interviewed for self than those who are employed. Receipt of unearned income, on the other hand, is not associated with with the likelihood of finding a person at home; thus, recipiency turnover for unearned income is not associated with a corresponding change in response status.

Conclusion

This paper has demonstrated the existence of some data quality problems in the 1979 Panel of the ISDP, at least when data are examined from more than one survey wave at a time. We have as yet no definitive explanation for these problems, but only a list of possible causes: edit, imputation, and processing procedures; matching difficulties; and response errors. It is likely, of course, that all contributed to the observed effects.

Although modelled in many ways on the 1979 Panel, the SIPP has adopted several modifications which may reduce the problem of heightened turnover in income recipiency between survey waves. First, the SIPP questionnaire includes procedures by which information brought forward from the previous interview can be verified and corrected, if necessary, at the time of interview. The identification and correction of incorrect information was not systematically addressed in the ISDP. Second, the SIPP exercises much

tighter control on the sample than did the ISDP, through an improved control numbering system, and improved check-in procedures in Census Regional Offices. These new procedures should help keep mismatches to a minimum in linking consecutive survey waves.

In the future, as SIPP data become available we will monitor them closely for evidence of the type of problem we have demonstrated here. In addition, we will seek to ensure that data which might help pinpoint the cause of the problem (for example, match certainty indicators and edit and imputation flags) are systematically gathered and maintained. We are also planning a more active program of investigation--a record check study matching selected SIPP income receipt and amount data with existing administrative records. Such a study will contribute greatly to our understanding of the quality of SIPP responses, and will provide valuable direction to the development of any ameliorative actions to improve the quality of the SIPP.

Technical Note on Significance Testing Procedures: The following assumptions guided procedures for testing the significance of the between-wave versus within-wave difference in turnover rates:

Suppose five observations have common variance σ^2 and common correlation ρ. Then

$$\text{the variance of the average of four} = \frac{4\sigma^2 + 12\rho\sigma^2}{16}$$

$$= \frac{\sigma^2}{4}(1+3\rho)$$

and

$$\text{the variance of the average of four - the fifth} = \sigma^2 + \sigma^2/4\ (1+3\rho) - 2\rho\sigma^2$$

$$= (5/4)\sigma^2(1-\rho).$$

In this illustrative example, the effect of positive covariance among the estimates is to reduce the variance below the sum of the variances of the two components. For the tests in Table 1, the variance of the difference was estimated by

$$Var_{diff} = 1/N\ [1/16\ (p_{12}(1-p_{12}) + p_{23}(1-p_{23}) + p_{45}(1-p_{45}) + p_{56}(1-p_{56})) + p_{34}(1-p_{34})]$$

where N = the number of adult sample persons in the two consecutive waves and $p_{i(i+1)}$ = the turnover rate for month-pair i and i+1

which ignores all covariances, and thus is likely to be conservative as compared to the illustrative example.

FOOTNOTES

1/ In fact, if the analysis indicated any consistent tendency, it was quite the opposite of that proposed by Czajka--less turnover in the month-pair which linked the two survey waves than in those within a single wave.

2/ Some suggestive evidence on the extent of this problem can be seen in the fact that about 20 percent of the entries in the Lepkowski and Kalton tables are of the "no match" variety, with data available for only one of the two waves. In fact, it was the frequency of this outcome which led Czajka to believe that the supposedly "no match" cases were actually "no receipt," since the code "occurs too often to reflect simply a failure to match records between waves" (Czajka, personal communication, 1983).

3/ Excluded from the tallies are the special subsamples of persons selected from lists of program participants, and persons who were not adult household members during both of the consecutive survey waves. Sample weights were not used for the tallies, and all analyses used the unweighted survey data.

4/ See the Technical Note regarding the procedures for significance testing.

5/ An explanation is in order regarding the last column of Table 1. In the design of the 1979 Panel, a randomly selected one-third of the sample was not administered a wave 4 interview, but skipped directly from wave 3 to wave 5. Thus, the first two sets of linked survey waves--1&2 and 2&3--contain the full respondent sample, sets 3&4 and 4&5 contain two-thirds of the sample, and set 3&5 contains one-third of the sample.

REFERENCES

Czajka, J. "Subannual Income Estimation." Paper 7 in M. David (ed.), Technical, Conceptual, and Administrative Lessons of the Income Survey Development Program (ISDP). New York: Social Science Research Council, 1983.

David, M. "Measuring Income and Program Participation." In M. David (ed.), Technical, Conceptual, and Administrative Lessons of the Income Survey Development Program (ISDP). New York: Social Science Research Council, 1983.

David, M. (ed.) Technical, Conceptual, and Administrative Lessons of the Income Survey Development Program (ISDP). New York: Social Science Research Council, 1983.

Goudreau, K., H. Oberheu, and D. Vaughan. "An Assessment of the Quality of Survey Reports of Income from the Aid to Families with Dependent Children (AFDC) Program." Journal of Business and Economic Statistics, 2 (2), 1984, pp. 179-186.

Lepkowski, J. and G. Kalton. "Investigations of Longitudinal Imputation Strategies for the ISDP Research Panels." Survey Research Center, University of Michigan, 1981.

Moore, J. "Program Participation Turnover in Waves I and II of the 1979 Income Survey Development Program (ISDP) Test Panel." Unpublished Bureau of the Census memorandum, July 26, 1983.

Ycas, M. and C. Lininger. "The Income Survey Development Program: Design Features and Initial Findings." Social Security Bulletin, November 1981, pp. 13-19.

TABLE 1: BETWEEN-WAVE (p_{34}) AND AVERAGE WITHIN-WAVE ($\bar{p}$) MONTH-TO-MONTH RECIPIENCY TURNOVER RATES, AND ANALYSIS OF DIFFERENCE OF RATES (p_{diff}), FOR SEVENTEEN INCOME SOURCES IN FIVE SETS OF LINKED SURVEY WAVES (1979 PANEL, ISDP)

	Linked Survey Waves														
	1 & 2			2 & 3			3 & 4			4 & 5			3 & 5[1]		
Monthly Turnover Rate:	p_{34}	$\bar{p}$	p_{diff}[2]	p_{34}	$\bar{p}$	p_{diff}[2]	p_{34}	$\bar{p}$	p_{diff}[2]	p_{34}	$\bar{p}$	p_{diff}[2]	p_{34}	$\bar{p}$	p_{diff}[2]
Income Type:															
Wage and salary	8.98%	2.99	**	10.10	3.41	**	10.22	3.58	**	7.51	3.01	**	9.03	1.51	**
Self employment and farm	2.99	0.22	**	3.08	0.29	**	2.85	0.26	**	1.99	0.34	**	3.18	0.32	**
Social Security	1.21	0.13	**	1.05	0.11	**	0.98	0.13	**	0.61	0.15	**	1.18	0.08	**
Federal SSI	0.46	0.04	**	0.36	0.04	**	0.20	0.04	*	0.16	0.03	*	0.26	0.02	*
Unemployment compensation	1.12	0.67	**	0.99	0.56	**	0.89	0.59	*	0.97	0.73	*	0.77	0.58	
Veterans benefits	0.31	0.03	**	0.24	0.03	**	0.16	0.02	*	0.16	0.03	*	0.12	0.02	
Workmans compensation	0.47	0.23	**	0.23	0.15		0.26	0.15		0.14	0.15	-	0.26	0.18	
AFDC	0.33	0.07	**	0.34	0.08	**	0.26	0.08	*	0.34	0.06	**	0.46	0.07	**
Child support	0.37	0.14	**	0.35	0.13	**	0.34	0.14	*	0.25	0.13	*	0.26	0.09	*
Employer or union pension	0.27	0.04	**	0.23	0.04	**	0.27	0.05	**	0.20	0.04	*	0.29	0.02	*
Educational benefits	0.25	0.27	-	0.149	0.153	-	0.20	0.14		0.16	0.17	-	0.07	0.16	-
BEOG	0.48	0.21	**	0.09	0.16	⊖	0.30	0.15	*	0.42	0.35		0.24	0.25	-
Food stamps	1.47	0.48	**	1.23	0.34	**	1.21	0.36	**	0.98	0.35	**	1.13	0.28	**
Rental income	0.56	0.04	**	0.76	0.04	**	0.29	0.07	**	0.44	0.08	**	0.14	0.01	*
Assist. from relatives, friends	0.75	0.37	**	0.63	0.15	**	0.40	0.09	**	0.32	0.13	*	0.53	0.15	*
Lump sum payments	1.38	0.95	**	1.48	1.01	**	1.25	0.80	**	1.00	0.93		1.16	1.08	
Incidental or casual earnings	0.74	0.31	**	0.97	0.38	**	1.13	0.43	**	0.93	0.45	**	0.94	0.45	*
Number of Cases	13,157			12,751			8,568			8,639			4,154		

1/ See text footnote 5.

2/ See the Technical Note for a description of significance testing procedures, results of which are symbolized as follows:

**	$p_{34} > \bar{p}$, $z > 3.3$	($p < .01$)
*	$p_{34} > \bar{p}$, $2.0 < z < 3.3$	($p < .05$)
[blank]	$p_{34} > \bar{p}$, $0 < z < 2.0$	(n.s.)
-	$p_{34} < \bar{p}$, $-2.0 < z < 0$	(n.s.)
⊖	$p_{34} < \bar{p}$, $z < -2.0$	($p < .05$)

TABLE 2: AVERAGE BETWEEN-WAVE (p_{34}) AND WITHIN-WAVE ($\bar{p}$) MONTHLY TURNOVER RATES, AND AVERAGE DIFFERENCE OF RATES (p_{diff}), BY RESPONDENT PATTERN IN CONSECUTIVE WAVES; AND COMPARISON OF DIFFERENCES FOR PROXY SITUATIONS VERSUS CONSISTENT SELF RESPONSE (1979 PANEL, ISDP)

	Average[1] Monthly Turnover Rates (p_{34} and $\bar{p}$) and Differences[2] (p_{diff})									Comparison[2] of p_{diff} for for Proxy Versus Self-Response Situations			
	Respondent Pattern in Consecutive Waves									p_{diff} (s-p and p-s) minus p_{diff} (s-s)		p_{diff} (p-p) minus p_{diff} (s-s)	
	Self-Self			Self-Proxy & Proxy-Self			Proxy-Proxy						
Income Type	Average p_{34} (a)	Average $\bar{p}$ (b)	Average p_{diff} (c)	Average p_{34} (d)	Average $\bar{p}$ (e)	Average p_{diff} (f)	Average p_{34} (g)	Average $\bar{p}$ (h)	Average p_{diff} (i)	Average effect (j)	Number of + differences (out of 5)[3] (k)	Average effect (m)	Number of + differences (out of 5)[3] (n)
Wage and salary	6.54	2.88	3.66	15.52	3.87	11.66	13.38	3.83	9.54	8.00	5	5.88	5
Self employment and farm	2.53	0.31	2.22	3.57	0.29	3.28	3.18	0.19	3.00	1.06	5	0.78	5
Social Security	1.03	0.12	0.90	1.22	0.12	1.10	0.73	0.09	0.64	0.20	4	-0.27	0
Federal SSI	0.32	0.04	0.29	0.24	0.03	0.21	0.20	0.02	0.17	-0.07	1	-0.11	1
Unemployment compensation	0.93	0.61	0.32	1.20	0.81	0.40	0.75	0.51	0.24	0.08	3	-0.08	2
Veterans benefits	0.20	0.03	0.18	0.21	0.03	0.19	0.13	0.00	0.13	0.01	4	-0.05	1
Workmans compensation	0.29	0.18	0.12	0.26	0.20	0.05	0.19	0.12	0.06	-0.07	2	-0.06	2
AFDC	0.41	0.08	0.33	0.30	0.06	0.24	0.11	0.01	0.09	-0.09	1	-0.25	0
Child support	0.43	0.18	0.25	0.14	0.02	0.12	0.04	0.01	0.03	-0.12	0	-0.21	0
Employer or union pension	0.31	0.03	0.27	0.23	0.01	0.22	0.05	0.04	0.01	-0.04	2	-0.26	0
Educational benefits	0.13	0.15	-0.03	0.31	0.28	0.02	0.19	0.17	0.02	0.05	3	0.04	3
BEOG	0.18	0.14	0.05	0.47	0.47	0.00	0.60	0.35	0.24	-0.05	1	0.19	5
Food stamps	1.37	0.45	0.92	1.02	0.25	0.76	0.73	0.12	0.61	-0.16	1	-0.31	1
Rental income	0.46	0.05	0.41	0.46	0.04	0.42	0.32	0.01	0.31	0.01	3	-0.10	1
Assist. from relatives, friends	0.65	0.22	0.43	0.39	0.15	0.24	0.16	0.03	0.13	-0.18	1	-0.29	0
Lump sum payments	1.41	1.10	0.32	1.02	0.68	0.34	0.83	0.65	0.19	0.02	3	-0.13	1
Incidental or casual earnings	0.93	0.41	0.52	0.91	0.33	0.58	1.00	0.38	0.61	0.06	2	0.09	4

1/ Average rates are computed as the sum of the rates derived from each set of consecutive waves divided by five.

2/ Minor discrepancies in some differences are due to rounding.

3/ Columns (k) and (n) provide evidence on the consistency of proxy involvement effects across the five sets of consecutive waves. Entries indicate the number of times the p_{diff} for the proxy situation exceeds the p_{diff} for the self-self pattern.

FINDINGS FROM THE STUDENT FOLLOW-UP INVESTIGATION OF THE 1979 INCOME SURVEY DEVELOPMENT PROGRAM

Anthony M. Roman
and
Diane V. O'Brien

Demographic Surveys Division
Bureau of the Census
Washington, D.C. 20233

For presentation at the annual meeting
of the American Statistical Association,
Philadelphia, Pa., August 13-16, 1984.

Findings from the Student Follow-up Investigation of the 1979 Income Survey Development Program

Anthony M. Roman
Diane V. O'Brien

I. Background

The Income Survey Development Program (ISDP) was a research and development program established in the mid-1970's by the Department of Health, Education and Welfare (HEW) in conjunction with the U.S. Census Bureau to prepare for the upcoming Survey of Income and Program Participation (SIPP). The SIPP is the new survey conducted by the Census Bureau designed to satisfy a wide variety of data needs concerning the economic situation of persons and families living in the United States. Data collection for the first SIPP survey, the 1984 Panel, began October 1983.

The major purposes of the ISDP were the same as the goals set out for the SIPP: to improve current estimates of income and income change; to extend the scope and precision of policy analyses for a wide range of Federal and State tax and social welfare programs, and to broadly assess the economic well-being of the population.[1]

The ISDP conducted four field tests. All were experimental in nature as different concepts, procedures, questionnaires and recall periods were tested. The 1979 Research Panel was the largest and most comprehensive research effort conducted by the ISDP.

The 1979 Panel was a nationwide household survey with a total sample of 11,800 households drawn from 130 Census primary sampling units (PSUs). Of this total, approximately 9300 cases were selected from an area sample and 2500 cases were drawn from list samples. Data collection began in February 1979 and ran through June 1980. One-third of the sample households were interviewed each month during the interview period. Information was obtained on household composition, labor force participation, various sources of money and nonmoney income, taxes, assets and liabilities, and other related topics.

The 1979 Panel included many controlled experiments which tested alternatives for basic survey design. The major tests conducted were: household versus individual questionnaire format; self versus proxy respondent rules; and 3-month versus 6-month respondent recall.

As part of the research effort to test respondent rules, one unresolved issue concerned proxy interviews taken for college students not living at their parents' address. In order to test the validity of information collected for this type of proxy interview, and experiment was conducted during the November and December interviews of the 1979 ISDP Panel. This experiment was called the Student Followup Investigation. This paper discusses the objectives, design, and field procedures used for the investigation, and some preliminary results of this experiment.

---

1 Research Triangle Institute. 1983. The 1979 ISDP Research Panel Documentation. NTIS, Washington, D.C.

II. Purpose

Respondent rules during the 1979 Research Panel were to conduct a personal interview for each adult household member 16 years or older. If a self-response interview could not be obtained, the procedure was to accept a proxy interview from another household member who was knowledgeable about the absent person. In this survey, as in other Census surveys, students were considered as members of their parents' households until they established a permanent residence elsewhere. Therefore, the usual procedure for students living away from home while attending school was to treat them as household members who were temporarily absent and obtain proxy interviews from other members of their parents' household.

The fourth interview questionnaire (Wave 4) used during the 1979 Panel contained a special set of questions concerning post secondary educational enrollment and expenses. Thus, this interview seemed especially appropriate for studying the quality of proxy interviews for students, as compared to the student's self interview.

In order to measure the accuracy of information taken from proxy interviews for students living away from home, the fourth interview was obtained first by proxy at the parents' household, and then a self interview was obtained from the student at the student's school residence. This self-response interview is referred to as the "Student Followup Interview."

There were two basic purposes for conducting the Student Followup experiment:

1) To obtain the most complete and accurate information possible for items in the Education Expenses section of the Wave 4 questionnaire (such as school enrollment, tuition, fees, and living expenses), and

2) To determine whether proxy respondents at the sampled address are able to provide reliable information on labor force participation, income, education expenses and enrollment for students living away from home. This experiment would be tested by comparing the information obtained from the proxy interview taken at the parents' home to the self-response interview taken at the student's school address.

III. Procedures

The fourth interview questionnaire used in November and December was designed to identify students living away from their usual residence while attending school.[2] Only students who were actually staying at their school residence (either a dormitory, fraternity, house, apartment, etc.) during the time of the November or December interview were eligible for followup.

---

2 The question "What is the address and telephone number of the place where ... lives while attending school?" was included in the Education Expenses section of the Wave 4 questionnaire. In addition, interviewers asked the student's parents for permission to contact the student at the school residence for the followup interview.

Census interviewers were instructed to call their regional office within 24 hours of a household interview in which a proxy interview had been administered for a student who was absent and living away at school. The interviewer would then provide the office with the student's name and school address.

Census regional offices were responsible for the control and assignment of the student followup cases. The rules for assigning the cases were essentially the same as the ISDP rules for movers. If the student's school address was within 50 miles of an ISDP PSU, the office assigned the case for interview. As soon as possible thereafter, an interviewer would visit the student for an interview. Regional offices were instructed to always employ a different interviewer for the student's interview in order to eliminate any interviewer bias. Additionally, interviewers were instructed to accept only self-response interviews at the student's school address; no proxy responses from roommates or friends were allowed.

IV. Field Results

The analytic universe for the study was the totality of students in the 4th Wave of the 1979 Panel who usually lived away from home and were attending post secondary schools.[3] There were 443 such students identified. Of these, 117 (26.4 percent) were not eligible for interview since the school residence was more than 50 miles from an ISDP PSU and 54 (12.2 percent) were not eligible because the student was staying at home during the time of the 4th Wave interview.

Of the 272 cases assigned, 202 Student Followup Interviews were obtained yielding a response rate of 74.3 percent. Of the 70 noninterviews, 6 were cases in which the parents refused permission for the interviewer to contact the student.

The major reason for the noninterviews was that many students were not staying at their school address (because of Thanksgiving, Christmas and semester breaks) by the time the interviewer received the followup assignment. Although interviewers were allowed until the first week of December to obtain the followup interviews for students identified in November and until the second week of January for students identified in December, many students remained on some type of break later into December and January. This proved to be an inappropriate time of year for conducting interviews with students at their school address. However, in the case of the 1979 Panel, we overlooked this factor in the survey design in order to conduct the experiment in conjunction with the Education Expenses questions, which were set beforehand for the Wave 4 interview.

---

3 Since the 1979 Panel interview design called for Wave 4 interviews to be administered for an abbreviated period of two months, only two-thirds of the 11,800 household sample were administered during the 4th wave interview. Thus, the Wave 4 sample size was approximately 8,100 households.

Recommendations for future studies involving student interviews at their school address include: 1) obtain the school address in a previous wave's interview, allowing interviewers more time to contact the student, and 2) conduct student interviews during months which do not contain long school holidays.

V. Preliminary Findings

A. Data Set Creation

The first task in analyzing these data was the creation of a data set of matched responses from the student followup questionnaire and the proxy questionnaire administered during Wave 4 of the ISDP. During the matching process, 35 students (17.3 percent) could not be matched to the Wave 4 ISDP File. Attempts to reconcile the mismatches were unsuccessful. In all but one instance, the most basic identifiers for these 35 students did not exist on the Wave 4 ISDP File. Due to the time elapsed from the initiation of this followup study to the creation of the analysis data set, it has been extremely difficult to find out why these mismatches have occurred. Future studies should be aware of these problems and prepare for them. Omitting the 35 mismatches resulted in a data set of 167 matched responses. These data are analyzed in this report. In all but 2 instances, the variables analyzed are direct responses to questions on the ISDP form (i.e., they are not in any way computed). The only exceptions are "usual hours worked per week at all jobs" and "total pay before deductions from all jobs last month". These 2 variables are computed by summing the response from each reported job.

B. Relationship of Student to Proxy Respondent

The relationship of the student to the respondent serving as his/her proxy can be determined in most cases through their relationships to the household reference person. The reference person is that household member who is stated as owning or renting the residence. Table 1 identifies these relationships.

Table 1:
Student/Proxy Relationships to Reference Person

	Proxy:				
student:	ref. person	spouse of ref. pers.	child of ref. pers.	other rel. of ref. per.	unknown
ref. pers.	-	0	0	2	1
child of ref. pers.	60	81	6	0	11
other rel. of ref. pers.	3	1	2	0	0
Total	63	82	8	2	12

In 141 cases (84.4 percent of the 167 matched cases), the proxy was a parent of the student. This follows the pattern one might expect by intuition.

C. Wage and Salary Comparisons

The ISDP questionnaire was divided into several sections. One section was designed to identify receipt of income types while other sections obtained amounts. Persons were asked a series of wage and salary questions if they indicated in the recipiency section of the questionnaire that they worked at a job or business. One wage and salary record was created containing responses to the set of wage and salary questions for each job named. Thus, if a student had only one employer, a wage and salary record should have been created with the student's responses while another wage and salary record should have been created with the proxy's responses. The reference period used in the ISDP was the previous 3 months, but the wage and salary records were created on a jobs basis. Therefore, a reported job could have been held at any time during the 3 month reference period. In examining the 167 matched cases of self and proxy responses, the following breakdown was observed:

83 had at least one self and one proxy wage and salary record
53 had neither a self nor a proxy wage and salary record
27 had a self wage and salary record but <u>no</u> proxy
4 had a proxy wage and salary record but <u>no</u> self

If one assumes that the self response is correct, then the proxy failed to identify a job held by the student in 27 cases (24.5 percent). This appears to be rather substantial and indicates that a potential source of underreporting of wages and salaries exists if proxy response is used. The 4 cases in which a proxy record exists while no self record exists may be interpreted as a potential source of misreporting wages and salaries under proxy response.

In attempting to analyze particular wage and salary questions of interest, several conditions must be kept in mind. While 83 matched cases exist with both a self and proxy wage and salary record, the number of cases available for making comparisons for any particular question may be less. There are two primary reasons for this: 1) one interview may have proceeded in a fashion which asked the question of interest and a response was coded, while the other interview proceeded in a fashion which did not ask the question (i.e., due to the various possible skip patterns within the questionnaire), and 2) even though the question of interest may have been asked during both interviews, one may have resulted in a valued response[4] while the other did not. Valued responses are important in evaluating the quality of data obtained in a survey. They indicate both knowledge by the respondent about the subject matter being investigated and willingness to cooperate in the survey.

With this in mind, the percentages of coded responses which were valued (i.e., given that a question was asked, the number of times it resulted in a valued response)are presented in Table 2. It is seen that for several wage and salary questions, it is more likely that a self respondent will give a valued response. This is particularly evident with the "usual hours worked per week" and "hourly rate of pay" variables. In all but one instance, when a valued response was not given, a "don't know" was the recorded response.

Table 3 presents the mean value of self responses for seven wage and salary variables for three particular categories:

1) the proxy could not identify that the student had a job (i.e., no proxy wage and salary record existed but a self wage and salary record did exist)

2) the self response was valued while the proxy response was not (e.g., the proxy most likely responded "don't know"), and

3) both the self and proxy responses were valued.

---

4 Throughout this report, the term valued response is used to imply any response with a legitimate value for the question asked. Valued responses do not include refusals, don't knows, or responses whose value is considered out of range or in some other manner erroneous.

This table demonstrates that a pattern appears to exist in which proxies best identify jobs at which students earn the most money or work the most hours. The smaller the earnings or hours worked, the more likely the proxy will either not be able to identify the job or not be able to answer detailed questions about the job.

Several points should be noted concerning Table 3. The usual hours worked per week may seem rather high for student jobs. This is due to the reference period for these questions extending back into the summer months. Therefore, summer jobs in which the student may have worked 40 or more hours per week will be included in these summaries. This also explains the decreases in total monthly pay from three months ago to last month. Also, it is impossible to compute total monthly pay by using the usual hours worked per week and regular hourly rate of pay. This is because the values presented in these tables are means and concern the student's primary job. One student's primary job may have been three months ago while another's may have been last month.

The final table, Table 4, presents comparisons of the self and proxy valued responses. It should be noted that the estimated variances used in computing these confidence intervals do not take into account any sample design effects. The reason is that this analysis is considered preliminary and will be used to decide if a more lengthy detailed analysis seems warranted. The net result is that the intervals in Table 4 should be considered conservative, while conclusions of significant differences should be considered liberal. Computation of design effects may add a small degree of accuracy to results from this study, but it should be noted for future studies, that increased emphasis on obtaining responses from all sample students and their proxies would greatly enhance the accuracy of results. Of the seven wage and salary variables analyzed, two showed a significant difference at the .05 level. These were usual hours worked per week and regular hourly rate of pay, both for the student's primary job. In both instances, the proxy gave the larger valued mean response. It is interesting to note that for usual hours worked per week at all jobs, the mean self and proxy responses are not significantly different. This raises the question of the proxy and student possibly identifying different jobs as being primary.

D. Education Expenditure Comparisons

All 167 matched cases had both a self and a proxy education expenditures record, but 61 of these records were unavailable for this preliminary analysis. This was due to a flaw discovered in the manner in which the Wave 4 ISDP data were processed. Only rekeying of the questionnaires could retreive the data and this was deemed unwarranted at the present time. Therefore, 106 matched records were available for analyzing education expenditures.

Table 2 again presents the percentage of responses which were valued. It is obvious that a valued response is much more likely from a self respondent than a proxy respondent. This seems understandable for all variables except "amount paid by family on tuition and fees" since the other variables involve expenditures most likely handled directly by the student. In every instance that a valued response was not given, a "don't know" was the recorded response.

Table 3 displays the mean value of self responses both when the proxy has a valued response and also when the proxy response is "don't know". Three of the four variables considered do not appear to differ substantially between these two categories. Only the amount paid by family on tuition and fees exhibits a rather large difference with the mean self response being greater if the proxy has a valued response. This is consistent with the wage and salary results in that the more expensive the tuition, the more the proxy is likely to know about the amount. It may also help explain why so many "don't knows" were given by proxies in response to this question. Perhaps when the amount of tuition is low, the student is more likely to be directly involved in its payment (e.g., the student may pay the tuition from support supplied by the parent).

Table 4 again presents results of comparisons of self and proxy valued responses. Two of the four variables showed a significant difference at the .05 level. They were academic credit hours taken this term and cost of course materials. In both instances, the mean proxy response was larger. The mean proxy response was also larger for amount paid on tuition and fees but with a large estimated variance, no significant difference was detected.

E. Other comparisons

Two additional areas were investigated in this preliminary analysis. The first was educational assistance. Of eight assistance categories, only two had enough reported cases to analyze questions concerning amounts received. These were: Basic Educational Opportunity Grants (31 cases) and Government Scholarships, Fellowships, Etc.(11 cases). The results of comparisons of self and proxy valued responses are shown in Table 4. No significant differences in mean amount received were found for any of the assistance variables.

The last area investigated was receipt of interest income. Reporting of interest was handled in the ISDP questionnaire in the same manner as wages and salaries. That is, a person was asked a series of questions regarding amounts of interest if they indicated receipt of interest income in the recipiency portion of the questionnaire. For the 167 matched cases, the following breakdown occurred:

104 cases had both a self and proxy report of receipt of interest

30 cases had neither a self nor a proxy report of receipt of interest

27 cases had a self report of receipt of interest but <u>no</u> proxy report

6 cases had a proxy report of receipt of interest but <u>no</u> self report

Assuming the self response is correct, the proxy failed to identify that the student would have interest income in 27 cases (20.6 percent). Although this appears to be a large problem, interest income is poorly reported for all people. For example, in the 104 cases in which both the self and proxy respondent reported receipt of interest earned on the student's own accounts, 61.0 percent of the coded self responses were "don't know" while 81.5 percent of the coded proxy responses were "don't know". Considering the question on interest earned on the student's shared accounts, 69.4 percent of the coded self responses and 80.0 percent of the coded proxy responses were "don't know". Obviously, it appears that the quality of interest data for students is suspect regardless of whether a self or proxy interview is conducted.

VI. Conclusions

The aim of this preliminary analysis was to examine the self and proxy student data in order to decide if a more extensive investigation (e.g., effects of accepting proxy responses on overrall survey estimates) seemed warranted. Any inferences drawn from these data should keep in mind that the estimated variances did not reflect any sample design effects and that the size of the data set is quite small. Indeed, most comparisons were

based on less than 100 observations. Still, this study is unique and although somewhat flawed in administration and implementation, it is possible to make certain general remarks. When valued responses are available from both the self and proxy interviews, the quality of the proxy responses appears to be generally quite good. Substantially more data would be needed to derive better estimates of the difference between self and proxy response and to narrow the confidence intervals around these estimates.

A problem that does appear to exist is in obtaining a valued proxy response. Quite often, a proxy cannot identify a particular source of student income (e.g., wages and salaries) and even if they can identify it, they are more likely to respond "don't know" to the particulars about that source. A trend does seem to exist that the larger the income or expense, the better the proxy response becomes. Still, this implies that by using proxy responses, the lower range of income or expense amounts are more likely missed.

Finally, the main issues involved in interviewing students away from home are the impact of accepting proxies on overall survey estimates and the differential costs involved in obtaining self responses. Since no cost data is available from this study, an estimate of the additional amount required in obtaining self responses cannot be computed. It may be possible to make some very general comments about the potential impact of accepting proxies on overall survey estimates. Students living away from home make up less than 3 percent of the overall ISDP sample. With this in mind and the fact that results from this study indicate that proxies are more likely to miss only the smaller expense and income amounts, it may appear unlikely that overall survey estimates will be strongly affected. Still, the limitations of the sample involved in this study must be considered in any statement of results. For instance, students living more than 50 miles from an ISDP PSU were omitted from consideration. Also, problems were encountered in matching students to proxies and in losing some survey data due to a processing flaw. The effect that these students could have had on results from this study is unknown. In concluding, further detailed investigation of this particular data set is not recommended due to the limitations in the size and composition of the sample. Future study may lead to stronger results but based upon this preliminary investigation, it is recommended that while the self-proxy student issue should not be forgotten, it should not occupy a high place on the SIPP research agenda.

Table 2: Percentage of coded cases which were valued
(Numbers in parentheses are amounts of coded cases)

Variable	percentage from self response	percentage from proxy response	difference
Wage and Salary			
Usual hours worked per week at primary job	98.7% (n=76)	76.4% (n=55)	22.3%
regular hourly rate of pay at primary job	100.0 (n=66)	73.3 n=75)	26.7
total pay before deductions from primary job 3 months ago	100.0 (n=60)	100.0 (n=51)	0.0
total pay before deductions from primary job 2 months ago	100.0 (n=61)	100.0 (n=50)	0.0
total pay before deductions from primary job last month	100.0 (n=61)	98.0 (n=50)	2.0
Education Expenditures			
academic credit hours taken this term	98.0 (n=101)	50.0 (n=102)	48.0
amount paid by family on tuition and fees this term	95.0 (n=100)	74.5 (n=102)	20.5
cost of course materials this term	100.0 (n=100)	46.1 (n=102)	53.9
amount of monthly rent and utilities	100.0 (n=30)	66.7 (n=33)	33.3

Table 3: Mean self response for several categories of respondents
(Numbers in parentheses are amounts of valued self responses)

Variable	Mean value of self responses when: proxy could not identify that student had a job	proxy did not give a valued response	both self and proxy had valued responses
Wage and Salary			
usual hours worked per week at primary job	22.30 hours (n=20)	21.11 hours (n=38)	35.60 hours (n=37)
regular hourly rate of pay at primary job	$3.17/hour (n=16)	$3.46/hour (n=18)	$3.39/hour (n=48)
total pay before deductions from primary job 3 months ago	$111.81 (n=16)	$246.48 (n=27)	$378.42 (n=33)
total pay before deductions from primary job 2 months ago	$32.19 (n=16)	$97.82 (n=28)	$138.33 (n=33)
total pay before deductions from primary job last month	$37.25 (n=16)	$74.83 (n=29)	$100.00 (n=32)
usual hours worked per week at all jobs	25.76 hours (n=21)	24.75 hours (n=37)	41.37 hours (n=38)
total pay before deductions from all jobs last month	$94.47 (n=17)	$131.10 (n=63)	$114.53 (n=51)
Education Expenditures			
academic credit hours taken this term	- -	14.88 hours (n=49)	14.80 hours (n=50)
amount paid by family on tuition and fees this term	- -	$469.25 (n=24)	$1004.10 (n=71)
cost of course materials this term	- -	$105.59 (n=54)	$98.75 (n=44)
amount of monthly rent and utilities	- -	$148.67 (n=9)	$131.14 (n=21)

Table 4: Summary of Comparisons

Variable	Mean Self Response	Mean Proxy Response	Difference	95% Confidence Limits[5] Lower	Upper
Wage and Salary					
usual hours worked per week at primary job	35.60 hrs	40.57 hrs	-4.97 hrs*	-9.03 hrs	-0.91 hrs
regular hourly rate of pay at primary job	$3.39/hr	$3.54/hr	-$.15/hr.*	-$.29/hr	-$.01/hr
total pay before deductions from primary job 3 months ago	$378.42	$336.09	$42.33	-$56.80	$141.46
total pay before deductions from primary job 2 months ago	$138.33	$121.52	$16.81	-$30.79	$64.41
total pay before deductions from primary job last month	$100.00	$106.56	-$ 6.56	-$22.56	$ 9.44
usual hours worked per week at all jobs	41.37 hrs	40.55 hrs	0.82 hrs	-3.55 hrs	5.19 hrs
total pay before deductions from all jobs last month	$114.53	$114.38	$.15	-$21.65	$21.35
Education Expenditures					
academic credit hours taken this term	14.80 hrs	16.00 hrs	-1.20 hrs*	-1.93 hrs	-0.47 hrs
amount paid by family on tuition and fees this term	$1004.10	$1157.63	-$153.53	-$591.87	$284.81
cost of course materials this term	$98.75	$120.84	-$22.09*	-$39.01	-$5.17
amount of monthly rent and utilities	$131.14	$121.38	$9.76	-$16.87	$36.39
Education Assistance					
BEOG assistance received 3 months ago	$299.61	$381.74	-$82.13	-$281.99	$117.33
BEOG assistance received 2 months ago	$194.48	$301.24	-106.76	-$302.28	$ 88.76
BEOG assistance received last month	$ 50.57	$ 14.71	$35.86	-$ 10.11	$ 81.83
Government scholarship assistance received 3 months ago	$ 87.55	$107.73	-$20.18	-$ 63.27	$ 22.91
Government scholarship assistance received 2 months ago	$154.75	$219.75	-$65.00	-$151.52	$ 21.52
Government scholarship assistance received last month	$ 8.00	$ 28.50	-$20.50	-$ 66.87	$ 25.87

* implies difference is significant at the .05 level

5 These limits are based on variance estimates which do not take sample design effects into account.

RESEARCH TRIANGLE INSTITUTE

THE ISDP 1979 RESEARCH PANEL AS A METHODOLOGICAL SURVEY: IMPLICATIONS FOR SUBSTANTIVE ANALYSIS

Richard A. Kulka
Research Triangle Institute

Paper prepared for presentation at the meetings of the American Statistical Association, Section on Survey Research Methods, Philadelphia, August 13-16, 1984.

Post Office Box 12194 Research Triangle Park, North Carolina 27709 Telephone: 919 541-6000

THE ISDP 1979 RESEARCH PANEL AS A METHODOLOGICAL SURVEY: IMPLICATIONS FOR SUBSTANTIVE ANALYSIS

The 1979 Research Panel was the third of three major field tests conducted by the Income Survey Development Program (ISDP) to fulfill its mandate to examine and resolve the technical and operational problems involved in the design and implementation of a survey mechanism that would meet the needs for improved data on income, assets, and program participation for program and policy analysis. By design, these field tests were of increasing size and complexity--progressing from a small experimental study of 2,800 households in five locations in 1977-78, and a national longitudinal survey of 2,400 households in 1978-79, to the 11,000 household 1979 Research Panel, conducted nationwide in 1979-80. They were also conceived as cumulative, with each generation of field tests offering incremental improvements in instrumentation and other survey design features, thereby moving closer to a prototype survey mechanism for use in the Survey of Income and Program Participation (SIPP), the large-scale national survey currently being conducted on a regular basis by the U.S. Bureau of the Census.

As the most proximal and realistic pilot or prototype for the SIPP, the 1979 Research Panel employed a longitudinal panel design, whereby persons at sample addresses were contacted early in the calendar year and recontacted at regular intervals (usually every three months) and asked about their income and other characteristics for the preceding few months. Drawing upon a detailed profile of income sources and labor force characteristics developed in the initial interview, it was possible in each succeeding interview to regularly update receipt and amounts of income for each succeeding quarter, along with any changes in household composition and labor force participation. A final annual "roundup" interview administered in the spring obtained year-end totals for selected income types to supplement information obtained on a quarterly basis, along with information on job benefits and taxes. In this way, a highly detailed record was built up for each person for the entire calendar year according to the schedule presented in Table 1. In addition, since less time was required to update basic information after the initial interview, time was available in later waves of interviewing to ask additional questions or questionnaire modules on: (a) topics of interest that were stable enough not to require updating on each visit; and (b) emerging issues of special interest to particular agencies or programs.

Table 1. 1979 Research Panel Interview Months and Reference Periods

Wave	Rotation group	Interview month	Reference period: 3 Months	Reference period: 6 Months
1	1	February 1979	Nov. 1, 1978-Jan. 31, 1979	-
	2	March 1979	Dec. 1, 1978-Feb. 28, 1979	-
	3	April 1979	Jan. 1, 1979-Mar. 31, 1979	-
2	1	May 1979	Feb. 1, 1979-Apr. 30, 1979	Nov. 1, 1978-Apr. 30, 1979
	2	June 1979	Mar. 1, 1979-May 31, 1979	Dec. 1, 1978-May 31, 1979
	3	July 1979	Apr. 1, 1979-Jun. 30, 1979	Jan. 1, 1979-Jun. 30, 1979
3	1	August 1979	May 1, 1979-July 31, 1979	-
	2	September 1979	Jun. 1, 1979-Aug. 31, 1979	-
	3	October 1979	Jul. 1, 1979-Sept. 30, 1979	-
4	1	November 1979	Aug. 1, 1979-Oct. 31, 1979	May 1, 1979-Oct. 31, 1979
	2	December 1979	Sept. 1, 1979-Nov. 30, 1979	Jun. 1, 1979-Nov. 30, 1979
5	3	January 1980	Oct. 1, 1979-Dec. 31, 1979	Jul. 1, 1979-Dec. 31, 1979
	1	February 1980	Nov. 1, 1979-Jun. 30, 1980	-
	2	March 1980	Dec. 1, 1979-Feb. 29, 1980	-
6	3	April 1980	Calendar Year 1979 (January 1, 1979 - December 31, 1979)	
	1	May 1980		
	2	June 1980		

Under this flexible "modular" questionnaire design, then, part of each interview was taken up by a set of "core" questions on income and other questions requiring quarterly update, while the remaining question sets--supplemental or "topical" modules--varied substantially from one interview to the next. As a result of the diligent application of this modular approach, the fully-implemented 1979 Research Panel produced an overwhelming array of data suitable for longitudinal analysis and a corresponding wealth of detailed socioeconomic data on more specialized issues suitable for complex cross-sectional analyses. The diversity of survey content of both types can be seen in Table 2, which summarizes all survey content by wave while also classifying question sets as either core sections or topical modules.

As the largest, most complex, and most "complete" of the ISDP field tests, the 1979 Research Panel is thus clearly of sufficient design, size, and interest to occupy social and policy analysts for many years to come, and its use for such substantive and policy research has been actively promoted (e.g., David, 1983; Kasprzyk, 1983a) coincident with the release of public use microdata files and technical documentation for these data by the National Technical Information Service (NTIS, 1983). It is important to keep in mind, however, that the fundamental purpose of the ISDP field tests--including the "state-of-the-art" 1979 Research Panel--was methodological: to serve as a flexible vehicle for conducting field experiments and feasibility tests to evaluate the effectiveness of alternative design features and data collection strategies (Ycas and Lininger, 1981). In this respect, too, the 1979 Research Panel was more comprehensive than the two previous field tests, serving to incorporate a wider range of explicit tests and experiments than the preceding surveys, while also serving as a crucible for the final shakedown and evaluation of a broad array of survey design features and implementation strategies to be employed in the SIPP. Reflecting its latter employment as a realistic dry run or dress rehearsal for the new income survey, virtually every feature of the 1979 Research Panel was subject to intensive methodological scrutiny, evaluation, and feasibility testing. Hence, the 1979 Research Panel is replete with such methodological assessments, and potential users of these data should be aware of their nature for at least two reasons. First, since most of these tests and experiments have not been fully analyzed or evaluated (some not at all), the potential for methodological as well as substantive analysis of these data is very great (cf. David, 1983). Second, since these methodological

Table 2. The 1979 ISDP Research Panel:
Overview of Survey Content by Topic and Wave of Interview

Topic or Module	Interview Wave: First: ISDP-101A	First: ISDP-101B	Second: ISDP-101	Third: ISDP-3101	Fourth: ISDP-4101	Fifth: ISDP-5101	Sixth: ISDP-6101
CORE SECTIONS							
1. Household Composition	ISDP-102	ISDP-102	--	--	--	--	--
2. Labor Force Participation	Sec. I	Sec. II	--	--	--	--	--
3. Income Profile							
3.1 Recipiency	Sec. I	Secs. I-II	--	--	--	--	--
3.2 Earnings and Employment	Sec. II	Sec. III	--	--	--	--	--
3.3 Amounts	Sec. III	Secs. IV-V	Sec. 3[1]	Sec. 3[2]	Sec. 3	Sec. 3[3]	--
4. Income Update							
4.1 Household composition	--	--	ISDP-102	ISDP-102	ISDP-102	ISDP-102	ISDP-102
4.2 Labor force participation	--	--	Sec. 1	Sec. 1	Sec. 1	Sec. 1	--
4.3 Income	--	--	Secs. 1-2	Secs. 1-2	Secs. 1-2	Secs. 1-2	--
5. "Missing Quarter" Information	--	--	Sec. 5	Sec. 5	Sec. 5	Sec. 5[4]	--
6. Children's Income and Transcription	Sec. IV	Sec. VI	Sec. 6	Sec. 6	Sec. 6	Sec. 6	Sec. E[5]
7. Adult Transcription	Sec. V	Sec. VII	Sec. 7	Sec. 7	Sec. 7	Sec. 7	Sec. F
8. Annual Income Roundup							
8.1 Earnings and Benefits	--	--	--	--	--	--	Sec. A
8.2 Asset Income	--	--	--	--	--	--	Sec. B
8.3 Taxes	--	--	--	--	--	--	Sec. C

[1]Included some additional questions on the value of selected assets.

[2]Included some additional questions on experience with and perceptions of selected programs.

[3]Included some additional questions on asset value and liabilities for businesses, farms, and other income-producing assets.

[4]Included some additional questions on multiple frame eligibility and other questions on sample coverage.

[5]Excluded questions on children's income.

Table 2--Continued

SUPPLEMENTAL OR TOPICAL MODULES							
1. Subjective Measurement of Economic Well-being							
1.1 Perceptions of income adequacy	Sec. VI	Sec. VIII	Sec. 8	Sec. 8	--	--	--
1.2 Evaluation of 1979 financial situation	--	--	--	--	--	--	Sec. G
2. Program Eligibility							
2.1 Relevant assets	--	--	Sec. 4	--	--	--	--
2.2 Deductible and offsetting expenses	--	--	Sec. 4	--	--	--	--
2.3 Disability status	--	--	Sec. 4	--	--	--	--
3. Attitudes of Program Eligibles	--	--	--	Sec. 4	--	--	--
4. Personal History							
4.1 Migration	--	--	--	Sec. 4	--	--	--
4.2 Educational attainment	--	--	--	Sec. 4	--	--	--
4.3 Work history	--	--	--	Sec. 4	--	--	--
4.4 Marital history	--	--	--	Sec. 4	--	--	--
4.5 SSI recipient characteristics	--	--	--	Sec. 4	--	--	--
5. Child Care							
5.1 Arrangements and costs	--	--	--	--	Sec. 4	--	--
5.2 School meals	--	--	--	--	Sec. 4	--	--
6. Educational Enrollment and Expenses							
6.1 Postsecondary education	--	--	--	--	Sec. 4	--	--
6.2 Secondary school tuition and meals	--	--	--	--	Sec. 4	--	--
7. Net Worth							
7.1 Housing tenure and vehicle ownership	--	--	--	--	--	Sec. 4	--
7.2 Liabilities	--	--	--	--	--	Sec. 4	--
8. Pension and Retirement Coverage	--	--	--	--	--	Sec. 4	--
9. Informal Assistance							
9.1 Inter-household transfers	--	--	--	--	--	--	Sec. D
9.2 Personal assistance	--	--	--	--	--	--	Sec. D
10. Participation in WIC	--	--	--	--	--	--	Sec. D
11. Foster Care and Childhood Disability	--	--	--	--	--	--	Sec. D
12. Telephone Service Coverage	--	--	--	--	--	--	Sec. D
13. Emergency Energy Assistance	--	--	--	--	--	--	Sec. D
14. English Language Competence	--	--	--	--	--	--	Sec. G

assessments are an integral part of the total survey design, they bear directly in some cases on the likely quality of the data for substantive analysis, including: (a) potential differences in the nature or quality of data collected under different experimental variations; (b) the confidence one can have in the accuracy of certain data; and (c) the extent to which data collected under different procedures can legitimately be conceptually merged for certain analyses rather than analyzed separately.

OVERVIEW OF TESTS AND EXPERIMENTS

With the latter especially in mind, I have summarized in Table 3 the full range of explicit methodological research embedded in the 1979 Research Panel. Although, as noted earlier, virtually every aspect of the 1979 Research Panel was subject to methodological review and evaluation, in particular, five formal, controlled experimental comparisons of alternative design or data collection strategies were systematically incorporated in the survey, and seven other procedures were explicitly included to provide a focused nonexperimental assessment of their feasibility for implementation in the SIPP. For each of these 12 tests and experiments a brief description is provided, along with its basic design, a capsule summary of results, and a note on its possible implications (if any) for substantive analysis. While this summary oversimplifies somewhat the sophistication and diversity of this methodological research, it does serve to illustrate both the vast potential of these data for methodological analysis and some possible road blocks these experiments throw in the way for substantive analysis.

Controlled Experiments

The first of five formal experiments included in the 1979 Research Panel compared two alternative questionnaire formats for measuring income recipiency, one using a "household screening approach" to determine receipt of various kinds of income and the other a more conventional person-by-person "individual" approach. It was hoped that the former approach would reduce the time needed to administer the questionnaire without a corresponding reduction in data quality. Preliminary analyses by Coder (1980) and Kaluzny (1981) indicated few differences between the two approaches in estimates of income recipiency rates by type, and only a slightly higher incidence of "don't knows" and "refusals" under the household screening approach, but the average savings per household was only about five minutes.

Table 3. The 1979 ISDP Research Panel:
Overview of Tests and Experiments

Test or Experiment	Description	Study Design	Capsule Summary of Results	Implications for Substantive Analysis
ITROLLED EXPERIMENTS				
Alternative Questionnaire Formats	Experimental comparison of a "household screening" versus an "individual screening" approach for measuring income recipiency.	Random assignment to households of alternative forms (one-third to household approach) during first wave of interviewing only.	Preliminary results suggest that household screening format did not offer significant improvement over the individual approach in either time or data quality.	Data on income recipiency gathered during Wave 1 confounded by any potential form effects.
Self-Respondent Versus Proxy Respondent Rules	Experimental comparison of rules maximizing self-response with those permitting proxy interviews for absent household members.	Random assignment of households to respondent rule alternatives (one-third to self-respondent) throughout all waves of interviewing.	Slightly higher costs and somewhat better data under self-response rules vs. higher interview rates under proxy treatments.	Data gathered in all interview waves subject to additional variation due to respondent rule treatments.
Three-Month Versus Six-Month Recall of Asset Income	Experimental comparison of length of recall period for property income.	Assignment of random half-samples to alternative reporting periods in Waves 1-5.	Little analysis done to date.	Data on amounts of asset or property income potentially subject to variation due to length of recall period.
Variable Recall of Monthly Information	"Staggered" interview design providing systematic comparison of monthly information reported using a 1- , 2-, and 3-month reference period.	Equal groups of households randomly assigned to a different month for each wave of interviewing.	Preliminary analyses reveal no consistent evidence of systematic recall effects.	Staggered design complicates calculation of monthly and calendar quarter estimates. May reduce recall bias in estimates of monthly income.
Seven-Point Versus Ten-Point Attitude Scale	Experimental comparison of reports of subjective economic well-being as assessed by two alternative attitude scales.	"Split-ballot" technique with a random half of self-respondents responding to each scale in Waves 1-3.	Ten-point scale resulted in broader distribution, reduced positive skew, and increased valid variance.	Measures of subjective economic well-being measures in Waves 1-3 subject to increased variance due to response scale differences.
ASIBILITY TESTS				
Student Follow-up Interviews	Examination of accuracy of information collected for students living away at school during interview period.	Administered Wave 4 questionnaire twice: by proxy at parents' address and in person at school address.	Preliminary analyses suggest quality of proxy response is reasonably good when proxies are able to report information.	Proxy data on college students somewhat incomplete, but not seriously so.
Nonfarm Self-Employment Income Test	Tested feasibility of using off-line mail-back surveys for obtaining quarterly estimates of self-employment income.	Respondents self-employed in a business or professional practice during Wave 1 mailed self-response mail questionnaire in first two calendar quarters.	Poor response rates forced abandonment of test after second calendar quarter.	Data on subannual self-employment inadequate for 1979 Research Panel.
Staggered-Interview Design	Study to determine whether increased interviewer experiences with the questionnaire and survey resulted in lower error rate.	Interviews conducted quarterly with one-third of the households interviewed during each of the three months comprising a quarter.	Little evidence that monthly interviewing results in substantially improved field performance.	See "4. Variable Recall of Monthly Information."

Table 3--Continued

Test or Experiment	Description	Study Design	Capsule Summary of Results	Implications for Substantive Analysis
9. Social Security Number Validation	Test feasibility of obtaining valid SSN's for sample household members.	Two rounds of validation, both including a computer match and manual search of administrative records, and field follow-up on invalid or missing SSN's.	Valid SSN's determined for 95.5 percent of cases included in project.	Survey information capable of being linked with administrative records systems using SSN identifier for high proportion of sample persons.
10. SSI Administrative Data Match	Match of survey and administrative records to validate information common to both sources and enhance survey data base.	Computer and clerical match of records using SSI subsample and SSI administrative tapes.	Final match rate of 99 percent achieved, but no analysis of data quality conducted to date.	None: list frame sample records not included on working files.
11. SSI Domain Match	Match to determine number of persons included in panel through area and BEOG subsamples also in SSI subsample frame.	Computer and clerical match of sample members with valid SSN's to SSI universe file.	Time consuming process resulting in "reasonable" match rate.	None: only area frame cases included in public use files. No need for multiple frame estimation.
12. Mover's Cost Study	Assessment of costs of following sample individuals moving to new address within 50 miles of an ISDP PSU.	Systematic recording by interviewers of their mileage and time spent in discovering, locating, and following up movers.	Mover household follow-up rate of 76 percent and a cost increase of about 8 percent attributable to following movers.	Substantial reduction in nonresponse due to loss of mobile sample members.

While differences between these two experimental groups are apparently small and undramatic, however, there is nevertheless little doubt that data on income recipiency gathered during Wave 1 of the 1979 Research Panel is confounded somewhat by any such potential "form effects". By definition the "household screening approach" compelled a higher level of "proxy" rather than self-reporting for income recipiency than under the individual approach (even under standard respondent rules), and because a few questions on income receipt that could not be covered on a household basis (e.g., verification of labor force, retirement, and disability status and Medicare and Medicaid coverage), and all questions concerning amounts of income received, were asked on a person-by-person basis even under the household screening approach, there is a distinct possibility that different types of income data may have been reported differentially under that approach but not when the individual approach was used. Moreover, although these alternative forms were used only during the first wave of interviewing, one cannot rule out the possibility of a longitudinal "forms effect" on these data. For example, serving as the "household respondent" in Wave 1 might result in more accurate reporting of one's own economic data across time than would be the case if each person reported income recipiency for themselves, and, conversely, those initially not serving as household respondents might subsequently be poorer reporters of income recipiency in Wave 2 when exposed to these questions for the first time.

The second experiment involved a controlled comparison of alternative respondent rules. Some households were interviewed using rules that required self-response except in special situations, while others were interviewed under "standard" respondent rules, whereby proxy interviews were accepted for absent persons from other household members when convenient. A number of different analyses have been conducted to date in an effort to study the effects of these proxy respondent rules and self-respondent rules on data quality, noninterview rates, and costs of data collection (e.g., Coder, 1980; Kaluzny, 1981, 1982; Kulka, 1983). In general, while the use of self-response rules results in approximately 20 percent more self-response (85 vs. 65 percent) and 4-6 percent higher interviewing costs than standard respondent rules, results on nonresponse and data quality are mixed. While the proxy treatment had a positive effect on household and person interview rates, self-respondent rules apparently resulted in somewhat better data (as implied,

for example, by the greater use of records, lower item nonresponse for certain key items, less rounding, and less variance in non-zero amounts), although some of these effects appeared to be somewhat smaller by Wave 2 (Kaluzny, 1982).

Unlike the "forms" experiment, the self-proxy response experiment continued throughout all waves of data collection, so that any potential influences on data quality or increases in the variance of key variables due to this experimental factor are likely to be found throughout the database. Yet, with few exceptions, the longitudinal implications of these alternative respondent rules have not yet been investigated. Changes in the proportions of proxy respondents or in the characteristics of proxy vs. self-responders under the two conditions may vary over time, thereby confounding somewhat longitudinal analyses of variables especially sensitive to respondent rules. More generally, suppose that a comparison of the 1979 Research Panel to other survey data suggested that the former provided more accurate data relative to an independent source of information. Suppose further that this improvement was directly attributable to the use of a maximum self-response rule (i.e., under regular proxy rules estimates were similar). Without making such an assessment, however, one might assume that <u>in general</u> the ISDP design results in better data and generalize this assumption to the SIPP, an erroneous presumption, of course, unless the SIPP were to employ rules maximizing self-response. Moreover, it is not difficult to see how such a methods "artifact" might similarly influence important relationships among variables of major policy significance.

The third experiment compared property or asset income amounts reported using a three-month reference period with that reported for a six-month recall period. The basic objective of this experiment was to determine if information on asset or property income data collected every six months would be as accurate as that collected quarterly. Results of this experiment would provide evidence on the magnitude of loss with the longer recall period (a critical ingredient in justifying the current four-month recall design for the SIPP), but very little of this analysis has been done to date (cf. Czajka, 1983).

The very reason for conducting this experiment, however, implies increased variation in reported amounts of asset or property income due to differences in length of recall period. Since the preceding three months are reported with an identical recall period by both groups every other wave, the influence

is not constant for all months. Thus, substantive analyses of a "common" three-month period may yield different results than that of a similar period where recall is three months longer for one group than the other. Similarly, quarter-to-quarter variation in asset income reporting may be greater within the six-month reporting group than within the three-month subsample. Moreover, since asset income recipiency is reported quarterly, the expected influences would likely be on asset income amounts, but a longer reporting period for "amounts" could also have an indirect adverse effect on reports of recipiency as well. In addition, if a recall effect of either type is present, such effects may either dissipate or increase in magnitude over the life of the panel (through Wave 5).

A fourth experiment, afforded by the use of a "staggered" interview design in which each quarter's interviewing was spread over three months with a variable three-month reference period (see Table 1), provided for a systematic comparison of income and other information reported for several months during the year using a one-, two-, or three-month reference period. Although the staggered design was not adopted for this reason, it provides a "natural" experimental design for the assessment of potential monthly recall bias by length of reporting period for virtually all income types and a wide variety of other variables. To date, however, only preliminary analyses of this natural recall experiment have been conducted (Kaluzny, 1981, 1982; Czajka, 1982), none of which have provided consistent evidence of a systematic recall effect.

Nevertheless, the implications of this monthly interview design for substantive analyses are considerable. On the positive side, the staggered interview procedure provides an ongoing measure of monthly recall bias, and to the extent that such bias exists, the varied recall period tends to minimize its effect (relative to more typical quarterly interviewing) when making comparisons of monthly changes, since income and other monthly data were always collected with the same average length of recall. On the other hand, the staggered approach introduces some substantial problems with regard to missing data and response variance for monthly and quarterly estimates. Point estimates for a given month are made with higher variation, and the staggered approach requires that calendar quarter estimates for two thirds of the sample be derived from data collected in two separate interviews, resulting in greater levels of missing data, linkage problems, and increased

month-to-month variation within quarters. For example, recent analyses of data from the 1979 Research Panel indicate a degree of variation in quarterly earnings greater than seems reasonable, and month-to-month changes in income recipiency generally tend to be greater between interviews than in the reference period reported within each interview (David, 1983:11; Moore and Kasprzyk, 1984).

A final area of controlled experimentation involved a comparison of reports by self-respondents of their subjective economic well-being assessed by using either a seven- or ten-point attitude scale. Preliminary analyses of these data (Vaughan and Lancaster, 1980) suggested that the ten-point scale achieved the desired effect of spreading out their distributions and reducing the positive skew associated with the seven-point scale, and additional analyses with a few variables suggested that this was primarily valid variance. In light of these findings, it would appear that these economic well-being measures may indeed vary significantly according to which response scale version is used, and substantive analysis involving these variables must take such variation into account, either by conducting separate analyses with these alternative measures or by directly including in the analysis a variable indicating which of these two experimental versions was used.

Feasibility Tests

In addition to the formal experimental comparison of self-respondent versus proxy respondent rules, two other more specialized respondent tests were carried out. One examined the accuracy of information collected for students living away at school during the interview period by administering the fourth wave questionnaire twice for absent students -- once by proxy at the parents' address and a second time in person at the school address. The basic objective of this study was to evaluate both differences in reporting and the additional burden imposed on field staff when students were followed to their temporary addresses. With regard to the latter, over one-fourth of the students identified lived at a school address outside the sample area, and of those assigned for follow-up, only 74 percent were interviewed, with most nonresponse due to inability to contact respondents at their school addresses. Preliminary analyses of data from the students interviewed indicate that when "amounts" or details are available from both the self and proxy interviews the

quality of proxy responses is generally quite good, but proxy respondents are frequently unable to provide a "valued response" at all (cf. Roman and O'Brien, 1984). In general, then, proxy data obtained for college students are clearly somewhat incomplete, but most analyses of data from the 1979 Research Panel should not be greatly influenced by these deficiencies, with the possible exception of those which rely on special subsamples containing a large proportion of college students and focus on variables especially prone to such proxy reporting error.

The second respondent test examined the feasibility of using off-line mail-back surveys for obtaining quarterly estimates of nonfarm self-employment income from respondents owning a business or professional practice. Because of poor response rates, this particular effort to measure subannual self-employment income was abandoned after the second quarter. Although some substantive analyses have been conducted using these data (e.g., Whiteman, 1983), methodological analysis took the form of additional experimentation with alternative procedures in an effort to improve this performance, none of which were very successful. The major implication of this feasibility test for social or policy analysts is that data on subannual self-employment income collected in the 1979 Research Panel are generally regarded as deficient.

The staggered interview design (mentioned earlier), which roughly tripled each interviewer's experience with a form, was itself a feasibility study. In addition to routine quality control interviews, an expanded reinterview program was initiated to determine whether such increased interviewer experience with the questionnaire and with the survey in general results in lower error rates. Research conducted to date provides little support for the proposition that monthly interviewing resulted in substantially improved field performance or data quality. Should such differences exist, however, this "interviewer experience" or "learning curve" factor may combine with potential recall factors to increase variation in these survey data, variation which would not be there if all interviews were conducted the first week of the calendar quarter, for example.

Two other feasibility tests incorporated in the 1979 Research Panel were designed to explore issues related to linkage of survey responses with data in administrative records systems. First, since the Social Security Number (SSN) is the identifier in most general use, a project to determine valid SSN's for sample households was conducted using two rounds of validation, both including

a computer match and manual search of Social Security Administration (SSA) administrative records. Through the use of those procedures and exploiting the panel design to obtain corrected SSN's from the field in later interview waves, valid SSN's were determined for 95.5 percent of the cases included in the project, a rate that might be improved with minor modifications in the future (Kaspryzk, 1983b). As a result, should access to administrative records systems be granted, substantive analysis using survey information linked to records data would be possible for a high proportion of persons sampled in the 1979 Research Panel.

Second, two distinct projects were undertaken to examine the feasibility of linking 1979 Research Panel data to benefit records of the Supplemental Security Income (SSI) program. The first involved a match of survey and administrative records using the 1979 Research Panel SSI subsample and SSI administrative tapes in order to validate information common to both sources and enhance the survey database. Overall, 3,950 sample persons in the 1979 Research Panel were matched with the SSI data sets, yielding a final match rate of 99 percent. However, analyses of data quality on this survey-administrative data match have not yet been conducted, and, since these list frame sample cases are not included in the public use microdata files (NTIS, 1983), this project is of no particular concern to substantive data analysts. Similarly, the second linking project, an SSI "domain match", was designed to determine the number of persons included in the panel through the area and Basic Educational Opportunity Grants (BEOG) subsamples who were also in the frame used to select the SSI subsample. Employing a match indicator code algorithm using validated SSN's, in combination with name, race, and date of birth, a reasonable match rate was achieved, albeit over a longer time period than would be required to support multiple frame estimation (Kaspryzk, 1983b). Since only the area frame cases are included in the public use files, however, multiple frame estimation is not required for substantive analyses of these data.

Finally, in an effort to determine the incremental costs of following movers (an integral feature of the survey design for the 1979 Research Panel and the SIPP), interviewers were asked to keep a systematic record of their mileage and time spent in discovering, locating, and following up persons or households that moved. A detailed analysis of this Mover's Cost Study is presented by White and Huang (1982), who (among other things) reported a mover

household follow-up rate of 76 percent (with an eligible person interview rate of 92 percent in interviewed households) and a cost increase of approximately 8 percent attributable to following movers. Of particular interest to potential policy analysts of these data is that nearly 78 percent of the 1979 Research Panel Wave 6 sample households had never moved. Relative to other longitudinal databases where movers are not followed, then, sample attrition due to this factor is clearly lower in the 1979 Research Panel, and estimates involving variables related to residential mobility less subject to such nonresponse bias.

CONCLUSION

In conclusion, from this brief overview it should be clear that, aside from their great analytic potential, some of these tests and experiments may also have a deletorious or confounding impact on certain substantive analyses that might be conducted using data from the 1979 Research Panel. And, in a few cases, these methodological have some positive implications for such analyses as well. Such implications range from some obvious deficiencies in some of these data highlighted by these field tests to more subtle influences on data quality and variances due to the experimental treatments imposed on the survey design. Especially with regard to the latter, the positive benefit of including such methodological tests in the survey design is that the potential influence of such factors on substantive results from this survey may be directly assessed in data analyses. It is important to note, however, that if these factors are not so examined their influence may lead to distorted or spurious conclusions. By describing some possible road blocks that these tests and experiments may throw in the way for substantive analysis and interpretation, this paper has sought to illustrate the need for both consumers and analysts of these data to keep their methodological nature clearly in mind and, where possible, to assess directly the potential influence of these factors on research results.

REFERENCES

Czajka, John L.
1982 "How complete was food stamp reporting in Wave II of the ISDP Panel." Manuscript. Washington, D.C.: Mathematica Policy Research, Inc.

Czajka, John L.
1983 "Subannual income estimation." Pp. 87-97 in Martin H. David (Ed.), Technical, Conceptul, and Administrative Lessons of the Income Survey Development Program (ISDP). New York: Social Science Research Council.

Coder, John F.
1980 "Some results from the 1979 Income Survey Development Program Research Panel." Pp. 540-545 in Proceedings of the Section on Survey Research Methods, American Statistical Association.

David, Martin H.
1983 "Measuring income and program participation." Pp. 3-20 in Martin H. David (Ed.), Technical, Conceptual, and Administrative Lessons of the Income Survey Development Program (ISDP). New York: Social Science Research Council.

Kaluzny, Richard
1982 Evaluation of Experimental Effects, 1979 Research Panel--Wave 1. Final Report. Princeton, NJ.: Mathematica Policy Research, Inc.

Kaluzny, Richard
1982 Evaluation of Experimental Effects, 1979 Research Panel--Wave 2. Draft Final Report. Princeton, NJ.: Mathematica Policy Research, Inc.

Kasprzyk, Daniel
1983a "Some research issues for the Survey of Income and Program Participation." Pp. 686-691 in Proceedings of the Section on Survey Research Methods, American Statistical Association.

Kaspryzk, Daniel
1983b "Social security number reporting, the use of administrative records, and the multiple frame design in the Income Survey Development Program." Pp. 123-141 in Martin H. David (Ed.), Technical, Conceptual, and Administrative Lessons of the Income Survey Development Program (ISDP). New York: Social Science Research Council.

Kulka, Richard A.
1983 "Tests and Experiments." Pp. 4-1 to 4-39 in P. Nileen Hunt, et al., ISDP 1979 Research Panel Documentation. Springfield, VA: National Technical Information Services.

REFERENCES (continued)

Moore, Jeffrey C., and Daniel Kasprzyk
1984 "Month-to-month income recipiency changes in the ISDP." Paper presented at the annual meetings of the American Statistical Association, Section on Survey Research Methods, Philadelphia, August 13-16.

National Technical Information Service (NTIS)
1983 Income Survey Development Program: 1979 Research Panel Documentation. Springfield, VA: NTIS.

Roman, Anthony M., and Diane V. O'Brien
1984 "Findings from the student follow-up investigation of the 1979 Income Survey Development Program." Paper presented at the annual meetings of the American Statistical Association, Section on Survey Research Methods, Philadelphia, August 13-16.

Vaughan, D.R., and C.G. Lancaster
1980 "Applying a cardinal measurment model to normative assessments of income: Synopsis of a preliminary look." Pp. 546-551 in Proceedings of the Section on Survey Research Methods, American Statistical Association.

White, Glenn D., Jr., and Hertz Huang
1982 "Mover follow-up costs for the Income Survey Development Program." Pp. 376-381 in Proceedings of the Section on Survey Research Methods, American Statistical Association.

Whiteman, T. Cameron
1983 "Lessons to be learned from ISDP: The measurement of nonfarm self-employment income." Pp. 73-82 in Martin H. David (Ed.), Technical, Conceptual, and Administrative Lessons of the Income Survey Development Program (ISDP). New York: Social Science Research Council.

Ycas, Martynas, and Charles A. Lininger
1981 "The Income Survey Development Program: Design features and initial findings." Social Security Bulletin 44 (November):13-19.

SOME DATA COLLECTION ISSUES FOR PANEL SURVEYS WITH APPLICATION TO THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

by

Anne C. Jean and Edith K. McArthur
U.S. Bureau of the Census

Prepared for presentation at
the 1984 Annual Meeting of the
American Statistical Association
in Philadelphia, PA
August 13-16, 1984.

Some Data Collection Issues for Panel Surveys with Application to the Survey of Income and Program Participation

Anne Jean and Edith McArthur
U.S. Bureau of the Census

Introduction: Recognition of a Need for a Longitudinal Survey.

The Survey of Income and Program Participation is designed to collect data which will improve our understanding of the income distribution, wealth, and poverty in this country. Information collected in the survey will be useful for planners and program administrators in areas such as income support programs and health care. The survey is longitudinal in the sense that the same persons are interviewed periodically over an approximately 2 1/2 year period. This implies following persons and updating information that reflects changes in their lives and in the composition of the households of which they are members--before, during, and after these changes occur. Persons in SIPP are interviewed every four months. At each interview, household members 15 years old or over are asked to report on income sources, amounts and employment for each of the previous four months.

With SIPP data we can now observe the effects over time of changes in receipt of different types of income upon the total income of a household; we can also see the effects of household composition change, such as the birth of a child or a marital separation, on participation in Federal transfer programs. In the past, analysts have often relied upon the income data collected in cross-sectional surveys, such as the March supplement of the Current Population Survey (CPS). The CPS describes household membership at a point in time, while obtaining income data for the entire previous calendar year. These data are consequently dependent on the household respondent's recall of events over the whole previous year. Thus many assumptions are made and monthly data cannot be collected accurately.

Implementation of a Longitudinal Survey.

The 1984 panel is the first panel of the SIPP. During the four months constituting Wave 1, that is October 1983 through January 1984, Census interviewers visited approximately 26,000 addresses located in 174 primary sampling units (PSUs) nationwide. The addresses were evenly distributed among four rotation groups, and each month one rotation group is assigned for interview. Nine interviews at four month intervals were scheduled for three rotations; the fourth rotation was scheduled for eight interviews (in the 1985 SIPP panel, and all future panels, all four rotations will have eight interviews scheduled).

The shift from an address sample for the first visit to a person sample in subsequent visits presented unique challenges to the planning staff, regional office staffs, and interviewers. Updating procedures for the address listings, noninterview classifications, interviewing procedures, and many other activities required for surveys maintaining an address sample were not appropriate. New controls and follow-up procedures, some requiring interregional office cooperation, were implemented. Interviewers received extensive training on new noninterview classifications and movers' procedures. Office staff maintained extensive clerical controls to guarantee the receipt of control cards and questionnaires from interviewers and to monitor the processing of over 40,000 person records that were uniquely identified.

The remainder of this paper describes the Wave 1 field procedures associated with the implementation of the address sample and the follow-up procedures for subsequent waves. Included is an explanation of the SIPP identification system and those field operations designed specifically for sample maintenance

and control. Some preliminary results of the 1984 panel follow-up are given and finally proposals for improving the follow-up system in future panels are discussed.

Wave 1 Address Sample Procedures

Field activities for the first SIPP interview were similar to operations undertaken for other major surveys that are basically cross-sectional, such as CPS and the National Crime Survey (NCS). Interviewers listed specific addresses of living quarters either prior to or at the time of the interview visit. Reasons for differences between the number of expected units based on census address lists and the number of units listed by the interviewer were researched by the office staffs. During the first interview, the address was verified, the unit was classified as a housing unit or OTHER unit according to census definitions. Coverage questions were asked to determine if EXTRA or additional units were located at the address, and the interview status of the address was recorded.

The interview status distinguished interviewed households from noninterviews. Noninterviews were further classified by type. For example, Type A noninterviews include all eligible households for which interviews were not obtained, such as refusals or cases where no one was home each time the interviewer visited. Type B and C noninterviews were recorded for addresses containing no eligible household such as vacant addresses, or houses under construction or being demolished.

In an interviewed household the interviewer listed all persons living or staying at the address, and applied a set of household membership rules to classify each person. Listed persons were classified as household members if the sample address was their usual place of residence as of the date of

interview. The specific rules for household membership in SIPP are identical to those used in CPS. All household members listed in Wave 1 were designated as sample persons. After listing all household members, demographic information, such as age, sex, and relationship, was obtained for each household member and a SIPP questionnaire was completed for each household member who was 15 years or over.

Development of Special Procedures for the Longitudinal Survey.

The procedural differences between SIPP and most other major household surveys conducted by the Census Bureau begin with the second interview. While other major surveys such as the CPS and NCS return to the same address for each subsequent visit regardless of whether the occupants of that address change, the SIPP interviewer returns to interview the same sample persons--that is, persons listed during the first interview. If persons move to a new address, they are followed and interviews are obtained at the new address. Between March 1981 and March 1982 almost 17 percent of the population of the United States moved.1/ If SIPP did not follow movers from the original sample households, we would lose the capability of observing the affects of many major changes in the original sample households, and the person sample would be biased, since it would not include movers.

Interviewers who discover that a Wave 1 sample person has moved (usually while updating the household roster) are instructed to inquire for new addresses at the original address and if future inquiry is necessary they are to contact mail carriers, rental agents, real estate companies and postal supervisors. Other sources may be used, such as an employer or a contact person (this is a person identified by the respondent during the initial interview as one who would usually know where the respondent was, such as a relative or a close personal friend). Occasionally, interviewers contact other persons with

---

1/ U.S. Bureau of the Census, "Geographical Mobility: March 1981 to March 1982." Current Population Reports, Series P-20, No. 384. Issued February 1984,

the same last name listed in local telephone books, although this procedure is not specified in their follow-up instructions. Beginning with the third interview visit, change of address notification forms are left with respondents and respondents are encouraged to mail these to the census regional offices (the address of the appropriate census regional office is preprinted on the form). In addition, advance letters are mailed to respondents before each interview; if the respondent no longer lives at that address the post office is requested to provide a forwarding address.

The regional office staff determines whether a new address will be assigned for a personal visit. Personal visits are required for all new addresses located in SIPP PSUs or within 100 miles of a SIPP PSU. Telephone interviews are encouraged for all sample persons who have moved to an address located more than 100 miles of a SIPP PSU within the United States. The following persons are excluded from mover follow-up:

(1) Persons who join Wave 1 sample persons in later waves are not followed to new addresses unless these additional persons remain with Wave 1 sample persons who are 15 years old or over;

(2) Persons who move out of the sample universe are not followed. These are persons who become institutionalized, move outside of the United States or live in an Armed Forces barracks;

(3) Children under 15 who move and are not accompanied by a sample person who is 15 years old or over are not followed;

The geographic area covering the personal visit follow-up is extensive. Based upon the 1980 census population distribution, about 130 million persons live in areas within SIPP PSUs; another 87 million persons live within 100 miles of the outer boundary of a SIPP primary sampling unit. We counted 226 million persons in 1980; 217 million are within our currently covered areas--96 percent of the population.

Of the 17 percent of the population that moved between March 1981 and March 1982, the great majority moved only a short distance--about two-thirds of the movers stayed in the same county (10 percent of the total population).2/ Thus, whether individual persons or even whole households moved, there is a good chance that if the new address could be ascertained the same interviewer would interview the sample persons at their new address if it was in the same county or in a nearby one. The remaining third who move outside of their original county are usually transferred to another interviewer. This occasionally involves a transfer between two census regional offices 3/.

Of the 26,024 addresses included in the original SIPP address sample, 19,878 addresses were interviewed households in Wave 1 and were reassigned for a second visit. The 6,146 addresses reported as noninterview at the time of the first visit were not reassigned. Of these noninterviews, 1,019 were eligible households who refused to participate in the survey, or were temporarily absent, unable to locate or not interviewed for other reasons. Survey planners were reluctant to reassign in Wave 2 those Wave 1 eligible noninterviews because of the added complexity for both the interviewers and the processing system 4/

Interviewers visited sample addresses for the second interview during February through May 1984 and attempted to locate and interview the approximately 40,000 sample persons interviewed during the first visit. New persons not present

---

2/ U.S. Bureau of the Census, op. cit.

3/ The United States is administratively divided into 12 geographic areas. Each area consists of a group of states under the jurisdiction of a census regional office.

4/ Wave 2 interviews for households not originally interviewed in Wave 1 require special procedures for constructing household rosters. For example, interviewers would need to obtain the names of persons living at the address as of a reference date four months prior to the Wave 2 interview. An appropriate Wave 1 person number would be assigned (see the SIPP Identification System explained later in this paper). However the 1984 computerized check-in system was designed to reject any Wave 1 person number that appeared for the first time in later waves.

initially were added to the household rosters, provided original sample persons were still included on the roster. Any new persons who were household members 15 years old or older were also eligible for interview. If no sample person remained at an address, no interviews were conducted at that address, but interviewers were required to follow the sample persons to their new addresses.

The SIPP Identification System.

The SIPP Identification System is a numbering system designed to provide a unique unchanging identifier for each person in an interviewed household. The person identifier is used to link data from more than one interview for the same individual regardless of what moves have taken place or what changes in household membership have occurred since Wave 1. In addition, the I.D. system provides the means for grouping individuals into unique households in each wave. This is an important attribute, which allows for the tracking and identification of changing household membership--persons moving away can be linked to each household of which they have been a member since their first interview. However, no attempt is made during the field operations to define or number each "different" household for longitudinal analysis.

The components of the operational SIPP identification system are:

PSU number - 3 digits
Segment number - 4 digits
Serial number - 2 digits
Address I.D. - 2 digits
Entry address I.D. - 2 digits
Person number - 3 digits

The PSU and segment numbers are assigned by Washington staff during sample selection. The 3-digit PSU number identifies a county or group of counties and is the same number used by other census surveys, such as the CPS and the NCS. As a sample of segments, that is, clusters of housing units, is drawn from a PSU, the segments are uniquely numbered within each PSU, using a 4-digit number. The clusters generally range in size from two to four housing units.

Office staff in the 12 regional offices are responsible for assigning the 2-digit serial number. The 2-digit serial number is assigned sequentially in Wave 1 to each SIPP living quarters within a segment.

The 9-digit combination PSU, segment, and serial number uniquely identifies each sample address for the first interview. As a result, SIPP households interviewed during Wave 1, (October 1983-January 1984) can be uniquely identified with these three components: PSU, segment, and serial number. The PSU, segment, and serial numbers never change, regardless of movers and new household formations.

For SIPP, a 2-digit address I.D. code is added by office staff to provide a means for identifying more than one unique household associated with the same PSU, segment, and serial number. This situation occurs after Wave 1, when an original Wave 1 household moves or splits up to form more than one household. The first digit of the address I.D. code indicates the wave a new address is first assigned for interview. The second digit sequentially numbers households originating from the same PSU, segment, and serial number. While not essential, an address I.D. code of 11 was assigned to all Wave 1 sample addresses. In later Waves, as SIPP sample persons move to new addresses, the office staff assigns new address I.D. codes to each new address brought into the survey by movers. The second digit of the address I.D. code is useful when sample persons who had been members of the same household during an earlier wave move to separate new addresses at a later wave. For example, the Smith's are a four person family; reference person, spouse and two children in their late teens. After Wave 1 the two children move out, one to work on a new power company project in another state and the other to take a job in yet another state. In Wave 2 the new address I.D. codes for the children's new addresses would be 21 and 22 respectively. The reference person and spouse who are still visited at their original address retain the original address I.D. code of 11.

Address I.D. codes assigned during a previous wave are deleted from the processing system for the current and successive waves if no SIPP sample persons remain at the address. Thus, the combination of PSU, segment, serial number, and address I.D. code uniquely identifies each sample address at each 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 identification number is a 5-digit number consisting of an entry address I.D. code and a person number. It is assigned by the interviewer as each person is initially listed on the household roster. As the interviewer lists the name of each person in the household, he/she transcribes the current 2-digit address I.D. code to each person's record. The 2-digit number is the entry address I.D. Next, the interviewer assigns a 3-digit person number to each person. Numbers 101, 102, and so on, are assigned to persons at the sample address in Wave 1; the numbers 201, 202, and so on, are assigned to persons added to the roster in Wave 2; and so forth. The first digit indicates the wave the person enters the survey. This 5-digit number consisting of entry address I.D. and person number is not changed or updated, except in rare instances of merged households which are described later.

Thus, the 14-digit combination of PSU, segment, serial, entry address I.D., and person number uniquely identifies each person in the SIPP survey and can be used to link data for persons across waves. The PSU, segment, serial, and address I.D. code uniquely identifies each household in a given wave; and the PSU, segment, and serial number can link all households in subsequent waves back to the original Wave 1 household.

As mentioned previously, the operational phase makes no attempt to apply longitudinal household definitions to the changing relationships, nor to number house-

holds longitudinally. However, as analysts develop longitudinal definitions, the current data base must be able to provide the information required to support these definitions. Further refinements in the questions asked at each interview may be implemented as the needs of a longitudinal household definition become more precisely specified.

The SIPP numbering system has several advantages over alternative schemes that have been considered:

(1) The portion of the control number consisting of PSU, segment, and serial number is similar to the numbering system used by other major surveys conducted by the Bureau.

(2) Interviewers are able to assign person numbers during the course of the interview. The person number is used in various parts of the questionnaire during the interview. This number is also transcribed to several other survey documents during the interview and immediately afterward during clerical coding operations. A person number assigned after the time of interview does not provide this immediate linkage.

(3) The person number itself has relatively few digits, reducing the possibility of transcription errors.

Several disadvantages have been noted:

(1) Duplications of person numbers for additional persons (persons added after Wave 1) can conceivably occur in situations where households have split and are in different regional office jurisdictions. The computer processing system identifies these duplications and the regional office staff corrects them during processing.

(2) Mergers between two separate sample households require special procedures. If this situation occurs, one set of controls is retained for the merged household. New person numbers are assigned to those

persons who lose their original identifiers. Interviewers record both the old and new I.D. numbers on the control card to provide a means for linking the two I.D.'s. By the end of the second Wave, this had occurred once.

Monthly Cross-Sectional Households

While the I.D. system provides identifiers for each household in a given wave, it does not identify households for a given month. Monthly cross-sectional households are not constructed in the field; rather they are constructed during processing using information obtained during each Wave. During each visit, demographic characteristics such as changes in marital status, changes in reference person (householder) status, and changes in household relationships are recorded on a control card. The same control card is used for each visit to the same address. If a sample person moves to a new address, the interviewer prepares a new control card for the new address and transcribes any information that is not expected to change. Date entered (month and day) and date left (month and day) are recorded on the control card for every entry and exit from an address. Reasons for entries and exits are coded:

Entry	1 - birth	)
	2 - marriage	)
	3 - other	)
	4 - 5/	)
Exit	5 - deceased	
	6 - institutionalized	
	7 - living in Armed Forces barracks	
	8 - moved outside of country	
	9 - separation or divorce	
	10 - person who joined household Wave 2 or later and is no longer living with any sample person	
	11 - other	
	99 - listed in error	

Date entered and left is used during processing to group persons into households for a given month. A person entering a household before mid-month is considered

---

5/ Code 4 is used in circumstances where a sample person moves to an address already occupied by persons not previously in SIPP. The persons not previously

to be a member for the entire month; a person entering after mid-month is considered not to be a household member for that month. A similar mid-month cut-off date is used for persons leaving households. As this monthly household determination is done during processing, it does not affect field operations, short of obtaining month and day of entries and exits.

Clerical Field Controls.

The SIPP movers' procedures have long been recognized as ambitious, requiring a system of field controls that are more extensive than those in effect for other major surveys conducted by the Bureau. Two standard forms are used for controlling interviewer assignments, and a third control was developed specifically for SIPP. All three forms are used during a clerical check-in at the regional offices.

An interviewer's Assignment and Control form is completed for each interviewer, listing every case in a given interviewer's assignment. A copy is sent to the interviewer and a control copy is kept in the office. As completed questionnaires are returned to the office, they are checked in against this form. A second control form lists all interviewers and the number of assigned cases for each interviewer. Tallies are kept as material is returned. This form gives supervisors a summary of the number of outstanding cases for a given month. The third control developed specifically for SIPP is a computer-generated listing of all persons listed as household members in Wave 1. It includes names, person numbers, interviewer codes, and interview status. The regional offices update the listing during each wave and account for every interviewed person as documents are received from interviewers. These three forms provide the basis of the clerical check-in and control. They must be updated to account for assignments that are transferred between interviewers and between regional offices, and they must be updated to include new persons entering SIPP after Wave 1.

Two other control forms are used by the offices to facilitate the movers' operation. One form is used to list the original address of a sample household along with all subsequent addresses. It is used primarily to control the assignment of address I.D. codes. A second form, a worksheet, is used for transferring cases from one interviewer to another by telephone. Because of time constraints, transfers are done by telephone; and required control card information--such as new address, names of persons, demographic information for the movers, etc.--must be obtained from the original interviewer and passed on to the new interviewer.

While the scope of this paper concerns field operations, some mention must be made of two major features in the computerized processing system designed for check-in and control.

(1) During the keying operation all persons listed on the control card who are 15 or over and are current household members must have an accompanying questionnaire. This check is done automatically at each keying station. Keying is done in the regional offices and immediate resolution of missing questionnaires is required.

(2) At the end of each of the four months of Wave 1, a centralized check-in is completed in Washington. A control card record must be transmitted for every person showing an active status on a master file maintained in Washington of all active records. Offices cannot close out an interview month until every active status person is accounted for and some demographic data--age, race and sex--is verified to make sure that we are not checking-in the wrong person. Each missing case is referred to the appropriate regional office for resolution.

Experience with Following Movers.

Available data for follow-up interviews conducted during February-May 1984 gives an initial indication of the success rates for the SIPP follow-up.

(1) Percentage of movers found: about 80%.

(2) Percentage of movers lost: about 20%--represents 0.9 percent of all eligible SIPP households.

When sample persons move from the address at which they were contacted in the previous Wave (four months before), interviewers are instructed to go through a series of steps to locate the new address. If all the steps are "Dead ends" they fill in a form which describes what they know about the mover situation for those sample persons. A review of the forms for Wave 2 available at the time this paper was written (they are submitted on a flow basis and a form was not submitted for some of the cases) illustrated the kinds of events that took place leading to the sample person's moving without leaving a trace. In about half of the cases all household members moved leaving no forwarding address. For another quarter of the cases one or more persons had left the household leaving other members behind but those other persons had no information about the departee's whereabouts. In an additional 15 percent of the cases, the spouse (usually the husband) left the rest of the family and the remaining spouse could not or would not give a forwarding address. The remaining cases showed a variety of events; for example, the person had moved out of his address and was maintaining no new household, rather he was just staying with various friends but the interviewer had no success in contacting him. The interviewers' comments showed considerable efforts in attempting to track these movers.

Recommendations for future SIPP Panels.

Improvements in the processing system and the expansion of follow-up procedures

are envisioned for future panels. These recommended changes are intended to improve sample coverage in a number of areas.

In the 1984 panel, persons who leave the sample universe--become institutionalized, leave the country, or live in an Armed Forces barracks--are dropped from the sample.6/ As of the 1980 census, about 2.5 million were currently inmates of institutions such as mental hospitals, homes for the aged and correctional institutions. Another 613,000 persons were living in military barracks. Demographers estimate that about 160,000 persons emigrate from the United States each year.7/ As average stays in nursing homes are less than 60 days and live discharges account for about 75 percent of the discharges, a sample person who goes into a nursing home is likely to come out before the end of the SIPP panel. According to current procedures, members of each of these groups are reinstated only if they rejoin a SIPP household.

For the SIPP panel beginning in January 1985, planning is underway to track sample persons who become institutionalized. Interviewers will obtain the name of the institution in which the person is residing. At each subsequent interview they will determine whether the person is still there and if the person has been discharged they will obtain a new address. It will then be possible to follow sample persons leaving institutions even if they do not rejoin active SIPP households. There are no current plans to track sample persons who move outside of the county or to an Armed Forces barracks.

Interviewers may return to an address in the 1984 panel and find that all original Wave 1 sample persons have left but one or more additional persons

---

6/ It was decided, not to obtain proxy information for sample persons (as well as other members of a household that has at least one resident sample person) who die while they are in a SIPP panel.

7/ Robert Warren and Jennifer Marks Peck, "Foreign-Born Emigration from the Unitd States: 1960 to 1970," Demography, Vol. 17, No. 1 (February 1980), pp. 71-84

(who joined households with sample persons after Wave 1) remain. In the 1984 panel no interviews are conducted at that address even though they lived with sample persons during at least part of the reference period, but interviewers still need to obtain a new address for the sample persons from the remaining household members. For future panels a final interview will be conducted for the additional persons. As in the 1984 panel, no subsequent follow-up is planned.

As described earlier, in the 1984 SIPP, only persons who are 15 or over are followed to new addresses; sample persons who are under 15 years old are not followed unless they move with a sample person who is 15 or over. However once they become 15 they are eligible for interview along with other members of their households. They are missed in the 1984 panel if they move before before turning 15 and are unaccompanied by a sample person who is 15 years old or older. Their absence may result in some bias in the survey data. In future SIPP panels, all sample persons who are 12 years old or older at the time of the first interview will be eligible for follow-up. When a person who was 12 years old at the time of the first interview moves by him- or herself to a new address, occupants of the new household will be interviewed according to standard procedures--that is persons 15 years old and over will be administered a questionnaire. When the sample person turns 15, that person will also be administered a questionnaire.

A number of other recommendations have been made for future SIPP panels. These include:

(1) Reassigning Wave 1 eligible noninterviews in Wave 2. Interviewers will be provided with instructions for obtaining household rosters and assigning person numbers retrospectively--i.e., as of a date approximately four months prior to the date of the second interview.

(2) Adjusting the computerized check-in system to allow for new serial numbers (representing persons or addresses) to be introduced in Wave 2. This will provide flexibility for including missed Wave 1 housing units.

(3) Developing a questionnaire that is appropriate for telephone interviews. This could be administered to persons who are not followed for a personal visit.

(4) While not strictly a recommendation, it is expected that increased automation over the next few years will eliminate much of the time consuming clerical operations associated with the check-in, control and monitoring of assignments.

Implementation of the recommendations depends primarily upon the feasibility of making the required changes in a complex processing system while adhering to the schedule for publication of data products.

In summary, SIPP has attempted an ambitious undertaking by implementing and attempting to improve an extensive follow-up program. Data users will be the ultimate beneficiaries and judges of the program's success.

MANAGING THE DATA FROM THE 1979 INCOME SURVEY DEVELOPMENT PROGRAM

by

Pat Doyle and Constance F. Citro

Prepared for the American Statistical Association Meetings
in Philadelphia
August 16, 1984

CONTENTS

Section		Page
I	OVERVIEW OF THE INPUT TO THE ISDP/RAMIS II DATA BASE.....	2
II	OVERVIEW OF THE ISDP/RAMIS II SYSTEM.....	11

FIGURES

Number		Page
1	Survey Waves -- 1979 ISDP Research Panel	3
2	Structure of the Reformatted Allocated Cross-Section Files from the 1979 ISDP Research Panel	6
3	File Structure for SIPPMASTER	12
4	RAMIS II File Reflecting Monthly Unit Composition (MH)	19

During 1979 and 1980 the Department of Health and Human Services and the Bureau of the Census, with support from other federal government agencies including the Food and Nutrition Service, USDA, administered a panel study of households representative of the civilian noninstitutionalized population in the United States called the 1979 Income Survey Development Program (ISDP) Research Panel. The survey was designed as the final pretest for the Survey of Income and Program Participation (SIPP) which had been under development since 1975 and was fully implemented in late 1983. The 1979 panel study was extremely complex due to the efforts put forth to improve the measurement of income, net worth, and program participation and to increase the information available on behavior, attitudes, expenses and disposable income of the population.

The complexity of the 1979 ISDP survey design led to the production of public use files which are cumbersome to use thus making it difficult to access the newly available data for research. The subject of this paper is to describe a project conducted by MPR under contract to the Food and Nutrition Service, USDA, to solve the data access problems through the use of data base management system technology. The DBMS chosen for this work was RAMIS II™ developed and distributed by Mathematica Products Group. The system developed by MPR is referred to as the ISDP/RAMIS II system.

In the subsequent section, an overview of the panel study with emphasis on the structure and contents of the publicly available files is provided. The report concludes with an overview of the newly created system with a summary of the data problems solved in the course of this work. For detailed information on the contents and use of the ISDP system, the reader is referred to Doyle and Citro (1984).

I. OVERVIEW OF THE INPUT TO THE ISDP/RAMIS II DATA BASE

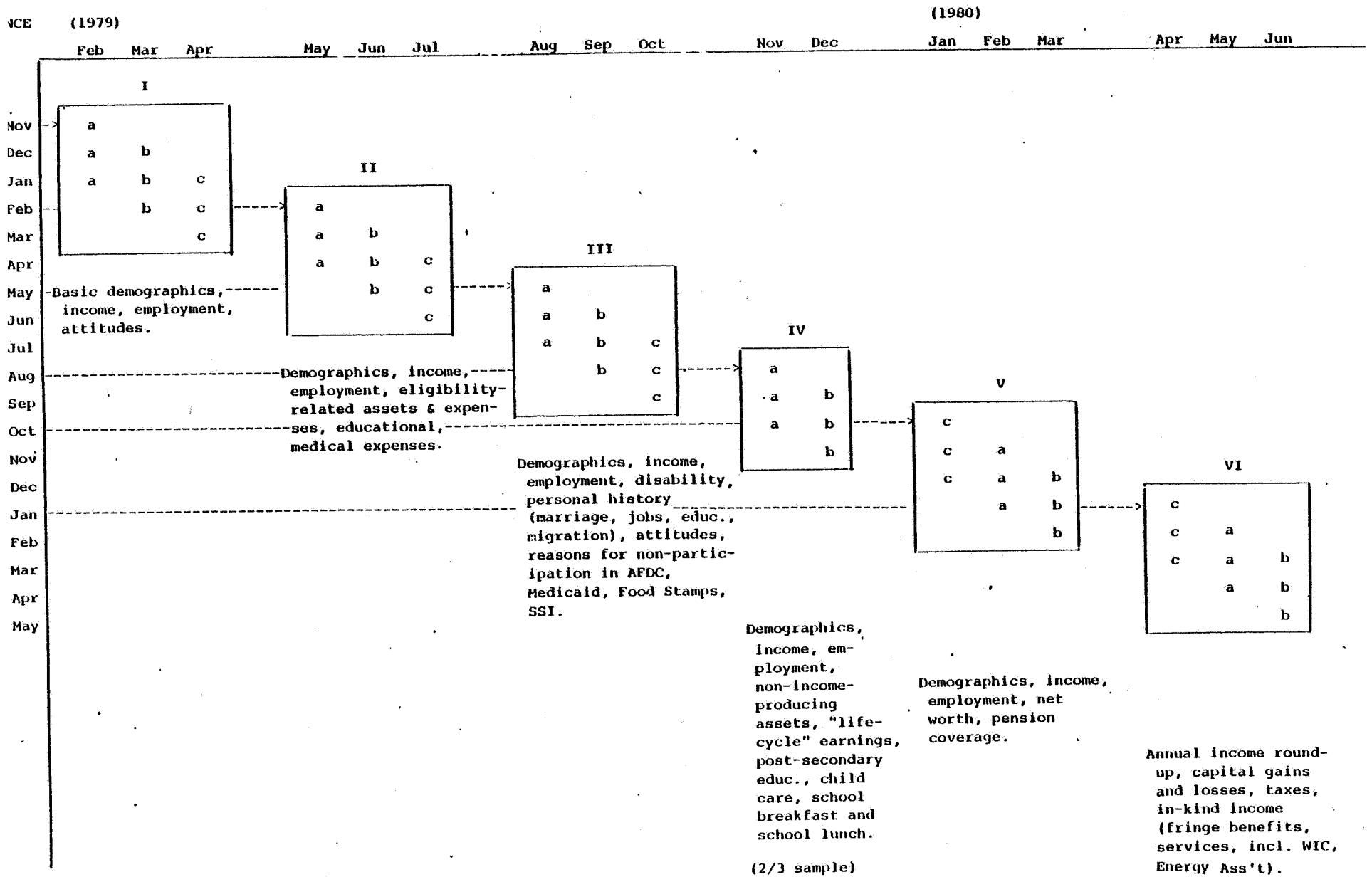
The 1979 ISDP Research Panel sample design provided for the inclusion of approximately 11,000 households drawn from a nationally representative sample frame and two supplemental frames extracted from administrative records. Of the 11,000 originally included in the sample, approximately 9,000 households were successfully interviewed in the first wave. Approximately 7,500 of those households were in the nationally representative sample, which formed the basis of the ISDP/RAMIS II system.

This survey consisted of six rounds (or waves) of interviewing; the first wave was fielded early in 1979 with subsequent interviews administered every three months thereafter. The Census Bureau employed a technique called staggered interviewing in the conduct of this survey. In other words, the Bureau divided the sample into three groups of equal size (called rotation groups) and administered the survey to one group each month. The first rotation group received the Wave 1 interview during February 1979 with subsequent waves occurring every three months thereafter. The second and third groups were administered the Wave 1 interview in March and April respectively. They too received subsequent interviews every three months thereafter. Figure 1 illustrates the outcome of this survey technique. One thing to note is that the third rotation group, although surveyed every three months, was not administered the Wave 4 instrument during the fourth round of interviewing. Instead they were administered the Wave 5 questionnaire with the Wave 6 survey following three months later. As a result, for one third of the sample there are five rather than six waves of data.

FIGURE 1

Survey Waves -- 1979 ISDP Research Panel

INTERVIEW MONTH



"a" = households in first panel

"b" = households in second panel

"c" = households in third panel

The instruments used in the 1979 ISDP Research Panel were modular in design with five core modules and a series of supplemental modules. The five core modules were administered during each of the first five waves of interviewing and covered the following topics:

- Household composition and characteristics at the time of the interview
- Person characteristics at the time of the interview and recipiency of income sources in the three months prior to the interview
- Earned income and employment status for each of the three months prior to the interview
- Unearned income for the three months prior to the interview
- Characteristics of children under 16 at the time of the interview.

Each of the supplemental modules was administered once during the survey, with Waves 2 through 5 containing at least one of the supplemental sets of questions. Figure 1 summarizes the contents of each of the five waves. As implied above, the reference period for most of the data collected was the three calendar months preceding the interview month. Hence, as can be seen from figure 1, the calendar period covered for each wave varies by rotation group.

Wave 6 was the final roundup interview and, in lieu of the core modules included in the first five waves, a series of questions designed to determine taxable income for 1979 were administered. Like the other waves, this last wave contained supplemental topical items as shown in figure 1.

Like the questionnaires themselves, the tape files available from the Census Bureau containing the results of this longitudinal survey are

organized in a series of six cross-sectional files, one for each wave. The structure of each is pictorially represented in figure 2. The data are grouped by household where a household represents a group of people residing together at the time of interview. Within each household there are records for each person present at the time of the interview as well as individuals who had been present sometime during the three calendar months preceding the month of interview. Associated with each adult there are a series of records which reflect the contents of the core modules described earlier. One or more family records are included if there were primary, sub or secondary families present at the time of the interview. There are no family records for primary or secondary individuals. For Waves 2 through 6 there are additional records containing the supplemental data collected during that wave. Generally, there is a separate record type for each topic covered.[1]

The contents of the files like the instruments include interview specific information as well as retrospective data. The retrospective data (with one exception) cover the three calendar months prior to the interview and are recorded in one of two ways.[2] First, most of the data vary by month and are labeled in relation to the interview month, i.e., "three months ago", "two months ago", and "last month". For the remaining retro-

---

[1]This disucssion of the tape files refer to those available to government agencies during 1982 and early 1983. In October 1983, new tape files became generally available to the public. These are similar to the tape files described here with one exception. For these publicly available tapes, the data were restructured to provide a complete representation of family groupings within households.

[2]The exception is asset income for one half of the sample. This is discussed further below.

FIGURE 2

STRUCTURE OF THE REFORMATTED ALLOCATED CROSS-SECTION FILES FROM THE 1979 ISDP RESEARCH PANEL[1]

Record Type		
02	Household Record	- One for each household in sample for the interview wave.
05	Person Record	- One for each adult and child in the household.
06	Wage and Salary Records	- Max. of 2 per adult.
07	Self-Employment Records	- Max. of 2 per adult.
08	Farm Record	- Present if adult had a farm or ranch.
09	Unearned Income Records	- Max. of 59 per person, each representing a different income type.
Varies	Supplemental Records	- Will be one or more depending on the interview wave; may be for the household or person.
19	"Family" Record	

[1]Reformatted allocated file is a term used to refer to the data file after it was resturctured from questionnaire image format and after imputations for item

spective data, the items refer to sometime during the reference period and only one summary item is included on the file.

There are a number of features of the ISDP survey and resulting public use files which complicate data access. These are summarized below.

- Staggered Interviewing. As discussed above, the Census Bureau used a staggered interview schedule to keep the field operations of the survey manageable. Thus, one-third of the sample was interviewed every month with the result that the reference period for the three months prior to each interview is a different calendar period for each third of the sample (called a rotation group). Hence, data from more than one interview must be accessed to study a common calendar period for the entire sample (except where the user can make do with the single calendar month that is common to all rotation groups within a wave).

- Skipping Wave 4. In order to schedule the Wave 5 interview so that asset balances as of the end of calendar 1979 could be obtained from all three rotation groups and also to schedule Wave 6 close to the time for filing tax returns, the interviewing schedule was altered to have the third rotation group skip over the Wave 4 interview. This meant that, although two-thirds of the sample cases have a full 15 months of data (from the five regular waves if they did not attrite), the other third has only 12 months. Moreover, the third rotation group does not have responses to any of the topical supplemental items asked at Wave 4, which includes information on school lunch and breakfast programs, child care arrangements, and educational expenses. It is also the case that, to analyze all cases for reference months 10 through 12, the user must access Wave 4 data for two rotation groups plus Wave 5 data for the third.

- Different reference periods for wave-specific information. At each interview, a number of variables, such as relationship to head of household, home ownership, and many others, were asked as of the time of the interview whereas data on employment and income were asked as of the three months preceding the month of interview. For any one interview, there is a potential mismatch between the wave-specific data and the monthly data, given that monthly data for the month of an interview were actually asked at the subsequent interview.

- Identifier problems. The Census Bureau encountered problems in uniquely identifying individuals across the survey waves, necessitating creation of a new unique person identifier, called the link index, as a separate file from the interview data files. It also turned out that the Bureau erroneously included some persons on the cross-section interview files who were not in fact present and vice versa.

- One-time wave-specific supplemental data. The fact that important data were asked on a supplemental one-time basis creates problems for using these items together with the monthly and quarterly data. For example, determining the extent of simultaneous participation in the food stamp, school lunch and breakfast, and WIC programs is rendered difficult because school lunch and breakfast participation was only ascertained at Wave 4 and WIC participation only at Wave 6.

- School lunch data problems. Examination of the data on Wave 4 by MPR revealed severe problems with the module containing information on day care and education expenses, and on participation in school feeding programs for children under 16. The Census Bureau included valid data only for the last child in a family, and erroneously wrote those data into the records for all other children.

- Lack of editing on Wave 6. In the case of Waves 1-5, the Census Bureau performed edits on demographic variables and also edited income recipiency flags. No editing was performed on the Wave 6 data, which were collected in an entirely different format.

- Asset income reporting experiment. The survey design included an experiment to study asset income reporting whereby half the sample reported asset income amounts for the prior three months at each interview (either monthly or quarterly amounts depending on the source), and the other half reported these amounts at every other interview for the prior six months. This creates practical problems of associating asset income data with other data for each month of the panel.

- Incomplete determination of monthly unit composition. Detailed information was obtained at each interview on household and family composition as of that point in time, and questions were also asked about arrival and departure dates for new and existing members. However, true monthly composition was not ascertained to go along with the month income and employment data. The design of the cross-section files, coupled with a high level of

noise in the data on arrival and departure dates, makes it very difficult to assemble a consistent stream of monthly unit composition indicators.

- Absence of longitudinal weights and imputations for missing data. The cross-section interview files contain weights and also imputations for missing income and employment data that were constructed strictly on a cross-section basis. For longitudinal studies, this is not a very satisfactory situation. For example, in the case of imputation, cross-sectional assignment of values for missing income observations may well distort patterns of receipt over time.

- Absence of longitudinal editing. With the exception of editing age and sex in the construction of the unique identifiers, no longitudinal edits were performed on the demographic variables.

These characteristics of the ISDP data files make retrieval of the information for analysis cumbersome and expensive. This is particularly true for longitudinal applications of the data such as the study of turnover in the Food Stamp Program. In order to facilitate data access, especially in the area of constructing longtudinal households, the Food and Nutrition Service, USDA asked MPR to develop a new system of data bases.

For purposes of this project, the original input file was a concatenation of cross-section files from all fives waves. The format was similar to that described earlier except that an additional control record was added. The records from all five waves were grouped by PSUSERIAL and a level 1 record was created which recorded information common to each group such as rotation.[1] In addition to inserting the level 1 record, the Bureau

---

[1]On the publicly available data bases, PSUSERIAL is a nine character field which uniquely identifies all households in Wave 1. Together with person number it was originally intended to uniquely identify persons followed in the panel. These nine charcters originally identifed the Primary Sampling Unit and Serial Number but they were subsequently scrambled for reasons of confidentiality.

also merged the link index (constructed unique person identifier) and longitudinal edited values of age and sex to this file. However, the Bureau deleted from this file the results of the cross-sectional imputations for income and employment data. The rationale for this omission was the unsuitability of these imputations for longitudinal analysis, the purpose of the concatenated file.

II. OVERVIEW OF THE ISDP/RAMIS II SYSTEM

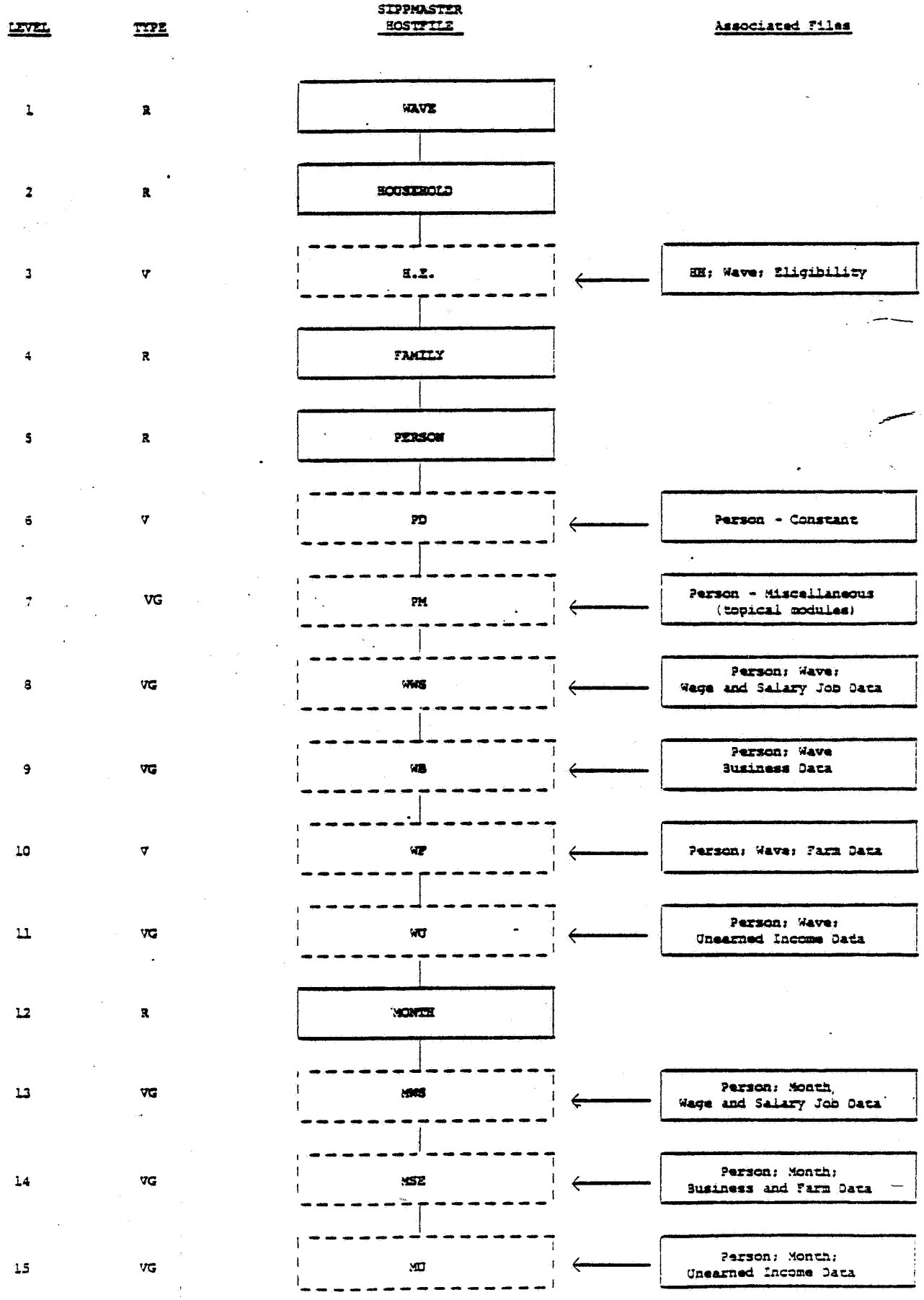
The objective of this data base development effort, as noted above, was to take the information available on the series of cross-section files described above and array it in a manner that would facilitate longitudinal as well as cross-sectional analysis. The results of this effort were two RAMIS II data bases, one called SIPPMASTER and one called MH for monthly households. SIPPMASTER is the main file in that all of the data collected during each wave are stored there. This file is used for all cross-section applications as well as longitudinal applications which do not involve the formation of longitudinal households or other groupings of individuals. The MH file is the data base designed to support the construction of longitudinal units. It essentially provides information on monthly household, family, and food stamp unit composition. The data in MH are arrayed to permit a user to develop a definition of longitudinality and apply that in the construction of a longitudinal unit file. Once the longitudinal unit itself is determined, the user can employ the data stored in SIPPMASTER to derive variables like total household monthly income which reflect the longitudinal unit characteristics.

The remainder of this section provides an overview of the contents of the ISDP/RAMIS II system. A detailed discussion of the motivation for choosing this file design and the procedures required to develop this data base is described in Doyle and Citro (1982).

SIPPMASTER. Figure 3 displays the logical organization of SIPPMASTER. It has a hierarchical structure with fifteen levels, five of which

FIGURE 3

FILE STRUCTURE FOR SIPPMASTER



Level Types:
R = Real

are real and ten of which are virtual.[1] The five real levels are wave, household, family, person and month. Some relevant comments on each of these levels follow:

Wave: (Level 1) Indicators for Waves 1 through 5 are contained in SIPPMASTER on level 1. The data from Wave 6 are treated as supplemental and are therefore stored in the virtual level PM (level 7). SIPPMASTER is physically separated into 5 data bases, one for each wave. They are linked together with RAMIS II USE commands to logically form one data base.

Household: (Level 2) This reflects household composition at the time of the interview. The household identifier (HHID) uniquely identifies households within wave. It cannot be used to identify households longitudinally. Non-interview households in each wave have entries at this level, however data for all other levels are zero. The contents of the household level consist of the data found in the household record in the cross-section files prepared by the Census Bureau.

Family: (Level 4) The family level simply identifies family units within households as they existed at the time of each interview. Primary individuals, secondary individuals, and out movers are treated as one person families.

Person: (Level 5) This contains interview specific data for each individual and retrospective data that were not collected for specific calendar months such as total weeks unemployed. The identifier for level 5 is the link index (called PERID in RAMIS II) so that each person sampled is identified in the same way across all waves. The data for the person level were derived from record type 5 of the cross-sction files. Some relevant points: out movers in a given wave are included for that wave but have 0 in the weight

---

[1] A virtual level is a level for which the data are not physically stored in the file. Instead there is an internal record of the location of another file which contains the information. With a DBMS, this other (or associated) file is accessed automatically when data from it are requested.

fields; the weights are cross-sectional; all person identifiers with values exceeding 200000 should be deleted for longitudinal analysis but not for cross sectional analysis; corrected age (CORAGE) should be used instead of edited age (AGEED) except that corrected age is 0 on Wave 2: income recipiency flags on level 5 are not to be used to determine item nonresponse as they were retained here for other reasons (for example, if the interest flag in Wave 1 is 1 on level 5 but there is no entry for that income type in the WU or MU associated files, then the person was reported to have had an interest producing asset but did not actually receive interest income during the Wave 1 reference period).

Month: (Level 12) This represents the reference period for each wave. However, all months in the survey have been numbered longitudinally so that, for example, the 3 months pertaining to Wave 2 are 4, 5, and 6. Aside from identifying the longitudinal reference months, this level contains numerous fillers intended to support the construction of longitudinal household (or other aggregate unit) files.

There is an important note to make regarding the month level and all associated files below it. In RAMIS II the months are numbered consecutively starting with "three months ago" information from Wave 1 and running through "last month" information from Wave 5. The association between these longitudinal reference months and the specific calendar month covered is a function of rotation group and wave. For rotation groups 1 and 2, the months range from 1 to 15 with three months assigned to level 12 for each wave. (For example, Wave 1 level 12 contains months 1, 2, 3 and Wave 2 level 12 contains months 4, 5, 6.) However, for rotation group 3, there are only 12 months due to the elimination of Wave 4 in the survey design. Since these are continuous in terms of calendar months covered, they are labeled 1 through 12. These conditions result in the following

situation for Wave 5: the three months assigned to level 12 for rotation groups 1 and 2 are 13, 14 and 15, but, for rotation group 3, the three months assigned are 10, 11 and 12.

The remaining data available through SIPPMASTER are stored in associated files which can be accessed directly if desired.[1] A summary of the contents of each follows:

HE: (Level 3) This contains the supplemental household data collection in Wave 2 (see figure 1). Note this level is only valid for households in Wave 2. The key to this level is HHID, the key to level 2 of SIPPMASTER.[2]

PD: (Level 6) This contains constant person demographics such as sex as well as an array of 16 zero/one flags denoting presence in the sample for each longitudinal reference month. This file is a single level file with one entry per person every included in the sample. The key is the link index (PERID).

PM: (Level 7) This is a virtual group when accessed from SIPPMASTER.[3] It is a two-level file containing supplemental data for each person. Level 1 identifies the individual using the link index (PERID) and level 2 identifies the supplemental record type. The variables are labeled with simple numeric names (INT1, INT2, etc.) since the contents vary by record type. There is a separate RAMIS II file called PMDESC which links each variable in PM to the original source by record type.

---

[1] Refer to footnote on page (14) for an explanation of associated files.

[2] Key fields are variables used by RAMIS II to search the data base for information. They essentially provide unique identifiers for each associated file.

[3] A virtual group is a hierarchical associated file where the identifier for the lowest level is not a key field in the main (or host) file.

WWS: (Level 8) This is a virtual group when accessed through SIPPMASTER. It is a three-level hierarchical file containing data specific to each wage and salary job reported in each wave (maximum of two) such as start and end dates or industry. The keys to this file are PERID, wave and job number. Note that jobs have not been linked longitudinally so that job 1 in Wave 1 is not necessarily the same as job 1 in Wave 2.

WB: (Level 9) This is a virtual group accessed through SIPPMASTER. It is similar to WWS except that it pertains to self employed non-farm employment.

WF: (Level 10) This is a two-level file identified by PERID and wave which contains wave specific data reported in the farm segment of the original survey instrument. It is not a virtual group when accessed through SIPPMASTER, because there is a maximum of one farm reported per wave.

WU: (Level 11) This is a virtual group when accessed through SIPPMASTER. This file contains unearned income-related items which do not vary by month such as type of deduction.[1] The identifiers are PERID, wave and income type. With one exception, income type corresponds to the codes used during data collection to identify each source of unearned income. Some unearned income data were collected together with farm-related information. These were loaded into WU with a previously unassigned income type code of 118. In the creation of WU, the observations in the 6-month recall asset experiment were assigned records in Waves 1 and 3 where asset income recipiency was indicated from the subsequent wave.

MWS: (Level 13) This is a three-level file which is a virtual group when referenced through SIPPMASTER. The keys are PERID, month and job number. The data stored in MWS are monthly data reported for each wage and salary job held

---

[1] The exception to this rule is that asset income reported quarterly or in 6 month sums was evenly allocated over months and stored in the MU file.

in the reference period of each wave. The job number corresponds to the job number discussed above for file WWS. However, the variable number is different and the complexity of the SIPPMASTER file design necessitates the use of RELATE when linking wave specific wage and salary data from WWS to the monthly data in MWS.[1] For this file, as is true for the other monthly files, there are two important notes to make. First, for months where earnings receipt and amounts were not reported and for which weeks worked was 0, there is no entry in the MWS file. Secondly, months are labeled as they are in SIPPMASTER (level 12), i.e., they are numbered consecutively beginning with the "three months ago" information from Wave 1. Furthermore, months 13, 14, 15 only exist for rotation groups 1 and 2 and the wave from which data pertaining to months 10, 11, 12 were derived varies by rotation group (see the discussion under month presented earlier).

MSE: (Level 14) This is a virtual group when accessed through SIPPMASTER with three levels, PERID, month and business number. It contains monthly data for self employment positions held as well as farm earnings allocated evenly across months. Self employment jobs are labeled 1 or 2 and farms are assigned a business number of 8. Refer to the MWS file for important notes regarding the creation of this monthly file.

MU: (Level 15) This is a three-level virtual group when accessed through SIPPMASTER. The levels are identified by PERID, month, and income type. As is true for MWS, the income type identifier here corresponds to that used in the WU file but RELATE is required to efficiently link the month specific and wave specific data relating to unearned income receipt. Furthermore, entries were created for this file only for the months when recipiency or amounts were reported and months were treated in the same manner described for MWS. Two additional notes are that the asset income reported for the 6-month recall experiment has been allocated to

---

[1] RELATE is a RAMIS II feature which gives that DBMS relational capabilities.

the proper month and that amounts originally reported as lump sums for the reference period have been evenly allocated across the relevant months. Income type 118 has been included in MU as was true for WU.

MH. Figure 4 describes the logical organization of MH. It is a relatively simple hierarchical file with five real levels and one virtual level. This file reflects the outcome of a complicated procedure designed to determine monthly household and food stamp unit composition from the data collected in the 1979 ISDP Research Panel. Documentation on the methodology employed in the development of this file is included in the (Doyle and Citro, 1984). The contents of this file are described below followed by a section summarizing how it is used to develop longitudinal units.

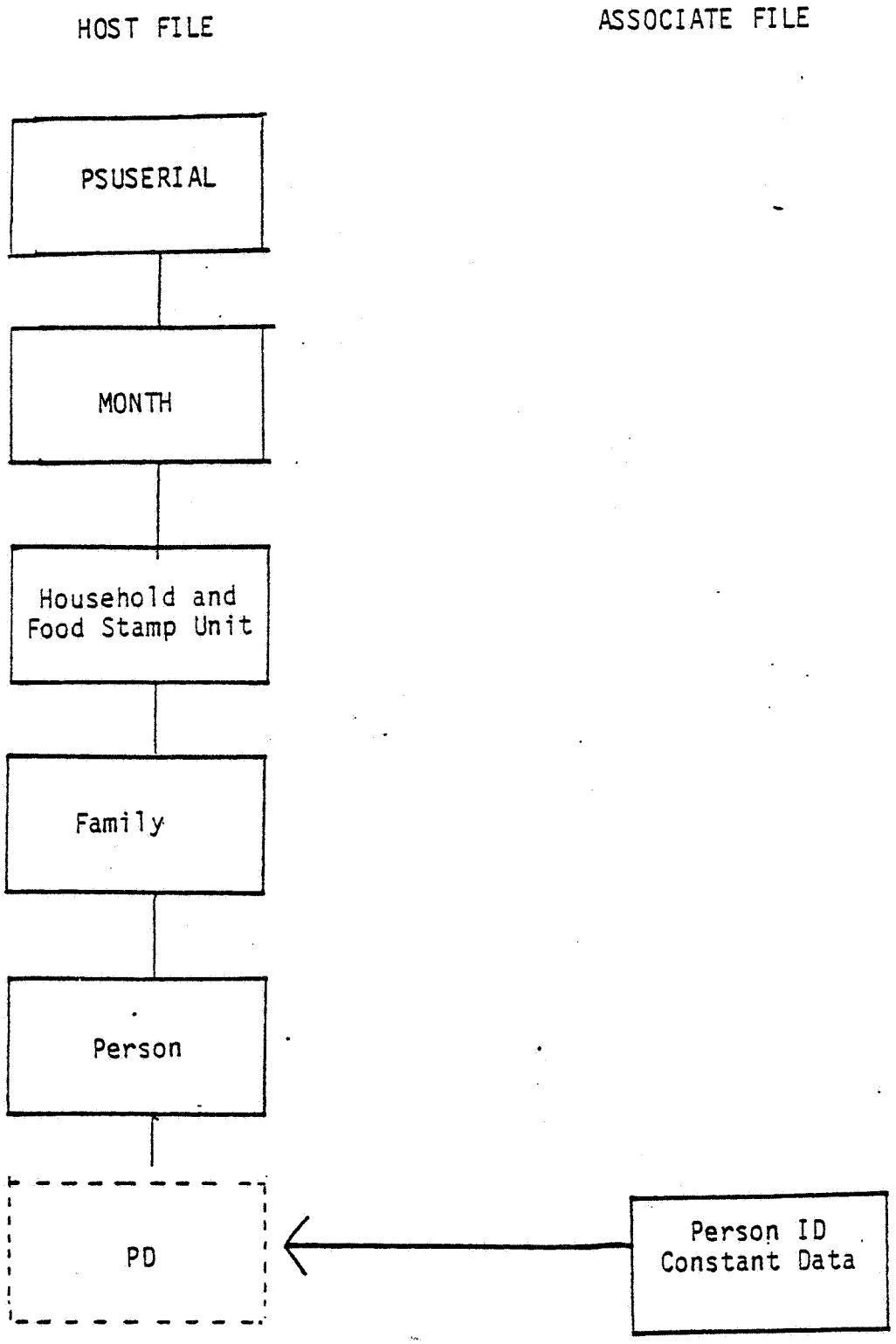
Unlike SIPPMASTER, MH contains a limited number of variables. It is comprised mostly of pointers detailing who lived with whom during each month covered by the first five waves of the survey. The remaining variables provide descriptive characteristics such as age and relationship to reference person which are necessary to effectively determine longitudinal units. Each of the levels of MH is described below.

PSUSERIAL: (Level 1) This level contains the scrambled values PSUSERIAL as well as the rotation group identifier. For the ISDP all persons who ever resided together have common values of PSUSERIAL, so this level was created to increase the efficiency of data retrieval and to minimize storage costs.

MONTH: (Level 2) This level simply identifies the month. Longitudinal reference months as described for SIPPMASTER were used. For rotation groups 1 and 2, the months range from 1 to 16 and for rotation group 3 they range from 1 to 13. Note that household composition can be

FIGURE 4

RAMIS II FILE REFLECTING MONTHLY
UNIT COMPOSITION (MH)



described for one more month than is covered by the retrospective data collected in the ISDP. This extra time period is the month of the final interview.

Household: (Level 3) This level describes who lived with whom during each month and the Food Stamp Program participation and benefits of that group. The contents are the monthly household identifier and food stamp recipiency and amount variables for up to two food stamp units.

Family: (Level 4) This is an indication of family groupings within monthly households. The contents are family identifier, family type, and family kind.

Person: (Level 5) This level contains an entry for each person for every month he or she was present in the sample. The key to this level is PERID, the same identifier used in SIPPMASTER. The other variables stored in this level are age, relationship to reference person, marital status, food stamp unit membership and variables necessary to link to SIPPMASTER.

PD: (Level 6) This is a virtual level in MH. The associated file is called PD and it is the same PD file accessed through level 6 of SIPPMASTER. It contains presence in sample indicators as well as constant demographic data such as sex.

The intended use of the MH data base is to determine longitudinal units. In developing the ISDP/RAMIS II system, one objective was to allow researchers flexibility in the development of the definition of what constitutes the same unit when viewed over time. For some applications it may be appropriate to define a unit as being the same from one month to the next if all adults remain the same. For another application it may be sufficient to only require that the reference person (household head) be the same. More complicated definitions may be required in other situations. An example might be that units are the same if the composition changes from one month to the next are restricted to birth of a child, loss

of a peripheral adult e.g., an older daughter leaves for college, or a death of one spouse in a husband/wife primary family.

Each of these three definitions can be specified with the ISDP/RAMIS II system as can many others. The procedure is as follows. Using the preferred definition, an algorithm for uniquely identifying each unit each month is developed. In the second example above, this would simply involve assigning the PERID of the reference person to the monthly unit as the identifier. Next, a comparison across months within PSUSERIAL groups is made. All monthly units with common values of the newly created identifier constitute one longitudinal unit. Finally, an extract is created which records the available information organized by the longitudinal unit identifier.

The available data from MH are primarily demographic, the exception being Food Stamp Program characteristics. The user will of course also desire economic data to support the anlaysis of the longitudinal units. This can be achieved through the use of SIPPMASTER in the following way. In addition to creating the MH extract described above, the user produces a second file with one record per person per month. The contents of this file are the SIPPMASTER keys (from level 5 of MH) and the longitudinal unit identifier. With this file a RAMIS II records management activity is used to load the longitudinal identifier into level 12 of SIPPMASTER. RAMIS II reporting capabilities can then be used to produce an extract containing the desired economic data. This file can be sorted by the longitudinal unit identifier and can then be merged with the demographic file initially produced from MH.

There are two points to make in the system described above. First, most applications of MH will be difficult to carry out with the nonprocedural report language RAMIS II users are familiar with. However, there is a procedural language interface called RPI which permits the use of the RAMIS II data base manager together with a procedureal language like FORTRAN. MPR has used this in the construction of an analysis file supporting the study of turnover in the Food Stamp Program (Carr, Doyle and Lubitz, 1983). Therefore MPR can either assist the ISDP/RAMIS II user in carrying out this activity or can develop an MH extract for users upon request.

The second point to make is that the ISDP/RAMIS II user need not have knowledge of the RAMIS II records management procedures to load the longitudinal unit identifiers into SIPPMASTER. MPR can provide a copy of existing procedures so that the user need only change the data set name of their MH extract and execute the programs.

BIBLIOGRAPHY

Carr, Timothy; Doyle, Pat; and Lubitz, Irene. "An Analysis of the Turnover in the Food Stamp Program." Draft Report (Washington, D.C.: Mathematica Policy Research, 1983).

Doyle, Pat and Citro, Constance F. "The ISDP/RAMIS II System and Its Development." Draft Report (Washington, D.C.: Mathematica Policy Research, 1984).

Doyle, Pat and Citro, Constance F. "Proposed Design Strategy for Storing and Accessing Data from the 1979 Income Survey Development Program Research Panel." Draft Report (Washington, D.C.: Mathematica Policy Research, 1982).